The Economics of Management Structures

Kieron J. Meagher

August 1997

A thesis submitted for the degree of Doctor of Philosophy of the Australian National University.

Unless otherwise indicated in the text, the work contained in this thesis is my own.

Kieron J. Meagher

Kieron Meagher

Acknowledgements

There are a great number of people who I should to thank here, and I apologize in advance to any whose names I forget to mention here.

My first thanks must go to my family, both for my physical existence and wellbeing, but also for providing such a lively intellectual environment. Special thanks to my brother Jonathan for the endless hours we have spent criticizing organizations. I hope that we can one day put this knowledge to more than academic use.

Secondly thanks Roy Radner and Tim Van Zandt for inventing the information processing approach to modeling organizations, and to John Tresler for bring it to my attention. Motivated by their research, I have found the formal study of organizations sufficiently fascinating to make the writing of this thesis more of a challenge than a trial.

Inspiration is one thing, work quite another. My greatest debt of thanks must go to my supervisors Simon Grant, Steve King and Gerry Garvey. It is they who have all, at one time or another, given their time to discussing my research and generally guiding my professional development.

While at ANU I have also benefited from the activities of a strong group of colleagues. Amongst these I would particularly like to thank Steve Dowrick, Lou Will, Cushla Paice, Adrian Pagan and all the students for their useful comments in seminars and on various drafts of parts of this thesis. My fellow students have provided a whole range of support far beyond academic concerns and I thank them most of all for their friendship. A couple deserve special mention: my fellow statistical theory reading group members Mardi Dungey and Vlad Pavlov, for everything from working through statistical problems to house moving and storage. Finally amongst my fellow students I would like to thank Mark Rogers, who is perhaps the perfect person with whom to share an office. He is also internationally competitive in the conversation and co-authoring areas.

Some faculty members require specific thanks for help on specific parts of this thesis. My empirical work would not have been possible without the help of Bruce Chapman, Jonathan Kelly, Mariah Evans and Jeff Borland. I would also like to thank Rohan Pitchford who has been a de facto fourth adviser.

The work on hierarchical management structures in a differentiated product market arose from conversations with John Miller and Scott Page. I am grateful to them for taking an interest in my work, inviting me to the Santa Fe Institute and laying the foundations for my work in computational economics.

My time in Canberra and at ANU has been rewarding. I would like to thank Bob Gregory for bringing me together with such a great dept (and Rae Carson and Stephanie Hancock for organizing the details). On the personal level I have benefited from having a great group of friends: Brendan, BJ, Nick, Bill, Adrian, Adam, Ash, Jamie and Attila.

It is typical to end by thanking ones spouse, without whom one could never have finished one's thesis. In my case this is not true, I am sure that I would have completed my thesis even if I had never met Karin Ronning. I thank Karin not for her help with my thesis (which has been considerable) but for my happiness.

Abstract

This thesis undertakes both a theoretical and an empirical analysis of management hierarchies within firms.

The theoretical work focuses on the information processing model of hierarchies pioneered by Radner and Van Zandt. Extensions are made to the one shot batch processing model to take account of both time based payment for labor services and the possibility of human error in determining the set of efficient hierarchies. It is shown somewhat surprisingly that changing the basis of labor payment does not change the efficient set of hierarchies. Two notions of fallibility are developed and applied in order to more finely characterize the set of efficient hierarchies.

The information processing model is also applied to a problem from Industrial Organization, namely choosing the characteristic for a differentiated product in the face of changing and imperfectly observable consumer preferences. It is shown that the optimal size of management structure is dependent on the market conditions that a firm faces.

The final part of the thesis takes the supervisor set of hierarchy models and tests their empirical implications for wages. It is found that both position in a hierarchy and the size of the hierarchy are significant determinants of wages within a human capital wage equation applied to Australian males.

Contents

Chapter 1

Introduction

1.1 Structure and Process

The best know observations on organization by an economist must surely be Adam Smith's description of a pin factory. He observes that one man, unfamiliar with the skills and methods of pin manufacture would hardly be capable of making twenty pins in a day. By comparison, ten men, organized in an integrated production process of eighteen trades produced upward of forty eight thousand pins a day. So that, through the division of labour, average output per person rose to four thousand eight hundred pins a day.

Adam Smith(1982,p112) suggests three channels through which coordinated organization gives rise to higher productivity:

> "This great increase in the quantity of work which, in consequence of the division of labour, the same number of people are capable of

performing, is owing to three different circumstances; first, to the increase of dexterity in every particular workman; secondly, to the saving of the time which is commonly lost in passing from one species of work to another; and lastly, to the invention of a great number of machines which facilitate and abridge labour, and enable one man to do the work of many."

The first channel which Smith identified is commonly known as specialization, and is the subject of much analysis in economics. The other two channels are examples of improved organization of the production process.

Production organization features prominently in the business press. The superior productivity of Japanese automobile manufacturers as a result of Just In Time inventory control, outsourcing and other so called modern manufacturing techniques, is one of the best known examples from recent history. Despite this widespread interest in production organization, recent economic advances in the theory of the firm, such as the principal-agent paradigm, and incomplete contracts emphasize incentives instead.

To understand exactly what I mean by production organization let us consider the example of Toyota[1]. Under the leadership of Ono Taichi, Toyota achieved higher productivity than the other major car manufacturers with initially lower levels of output. This was not achieved through investment in more

[1] A detailed history of the Toyota car company, including the results mentioned here, can be found in Cusumano(1985).

modern machines or by higher specialization but by more careful organization of the production process. As the following examples illustrate, on careful examination, many of the improvements in the production process were contrary to much received economic wisdom.

The introduction of the Kanban system reduces waste inventory in the production process by only producing inputs as they are required by a later stage. This entails smaller, more frequent production runs in the production of components. Smaller runs reduce the returns from economies of scale, but overall the Kanban system is more economical through a better coordinated production process, waste inventory is reduced.

The ideas on specialization and learning by doing are also turned on their head. Under Taichi the average number of machines operated by a worker increased from one to five. Although this change caused workers to have less experience with any one machine, output increased because more careful organization allowed workers to concentrate on only one machine at a time even though each worker had a number of machines running at the same time.

From this discussion of manufacturing production we learn that many factors affect productivity, and that one of the more important factors is the design and assignment of the tasks within the process.

Although manufacturing is of importance to economists the average economist is more interested in the general characteristics of manufacturing or production than in the details of the design of particular manufacturing processes. The details are seen as being better left to engineering, operations research or engineering economics. There is, however, a type of production present in all firms, the design of which is of great importance to economists. That production is the production of management. Management production typically involves the organization of labour (in what I call a management structure) to make decisions and to implement decisions through supervision,monitoring and authority.

The production process approach to management is one of the least studied areas in economics, and is the topic of this thesis.[2]

The traditional neoclassical approach to the theory of the firm has been to assume that an individual, the entrepreneur or manager, does all the managing. This simple approach has lead to a wealth of results. The economics of industrial organization has used this approach almost exclusively to analyze output, pricing and production decisions under different market conditions. The incentives literature has looked at the type of contractual relations between owner and managers (corporate governance), and owners and workers (principal-agent and incomplete contracts).

There have been a number of theories developed to explain the existence of management structures, which are also called administrative bureaucracies or hierarchies. It is something of a folk theorem that bounded rationality is necessary for a theory of management or administration. In the broadest sense

[2]The relationship between organization and decision making has not gone unnoticed. Cyert and March(1963) and Marschak and Radner(1972) being two classic works on the subject.

bounded rationality means there are limits on the ability of individuals to deal with information. The necessity of modeling bounded rationality and the interrelationships between individuals has made the modelling of management or administration difficult and has led, despite a great deal of interest, to a lack of research.

There are a number of types of bounded rationality which differ in the approach they imply (the taxonomy used here was originally proposed by Radner(1996)). One useful distinction is between resource-based bounded rationality and deciding-how-to-decide bounded rationality. Resource-based bounded rationality acknowledges that the acquisition or processing of information takes time and requires resources. The simplest example is incurring a cost in acquiring information. More significant applications are to monitoring and information processing.

Monitoring is essentially trying to pay attention to a number of signals available at the same time. Because an individual has limited attention she can only concentrate on one thing at a time and hence can only monitor one person at any given instant. Thus increasing the number of people supervised decreases the attention paid to any one supervisee. This problem can be mitigated by hiring other managers to share the task of monitoring, but these managers will in turn need to be monitored. This approach gives arise to a management structure for supervision and is the basis for the work in chapter 5.

Information processing bounded rationality assumes that decisions are made on the basis of information, that it takes time to process a piece of information and that it is important to make decisions in a timely manner. In this case, it may be possible to make a decision more quickly (and hence more profitably) by splitting it into a number of subtasks handled concurrently by a number of people. This gives rise to a management structure for making decisions. This approach, pioneered by Radner and Van Zandt, forms the basis for chapters 2-4.

The solution of resource-based bounded rationality problems can be handled by the application of the appropriate constrained optimization technique. This can be a technically demanding exercise but does not entail any major conceptual difficulties. Deciding-how-to-decide bounded rationality is a much deeper issue. If people are limited in their ability to deal with information, the models of the world they use may be simpler than reality (in some critical sense) and their decision making techniques less than fully optimal. This point of view is most strongly associated with the name of Herbert Simon. This issue seems vitally important to many areas of economics, especially in the theory of organizations. No widely accepted, operational solution has yet been proposed and further consideration of this problem lies beyond the scope of this thesis.

1.2 Organization of the Thesis

This thesis consists of four papers, chapters 2 to 5, and a concluding chapter 6. In these papers the subject of management structures is considered from three different perspectives. In chapter 2, we examine efficient information processing structures in an abstract setting without regard to the details of factor and product markets for the firm. The relationships between management structures and these two markets are considered in the other papers.

We discuss, in chapter 2, the Radner Van Zandt model of information processing which is the basis for the rest of that chapter, and for chapters 3 and 4. The first half of chapter 2 presents an improvement to the representation of labour cost in the information processing model so that it reflects the total amount of time spent working rather than just the number of people working. Using this new measure of labour cost, it is shown, somewhat surprisingly, that the set of efficient hierarchies remains the same.

In the second half of chapter 2, we attempt to incorporate ideas about human fallibility into the information processing model. Two types of fallibility are considered: errors which produce the wrong conclusion and errors that produce extra delay in decision making. Both types of fallibility enter into the firm's loss function in a lexicographic fashion and are used to give a sharper characterization of efficient hierarchies.

In chapters 3 and 4 we apply information processing hierarchies to a concrete problem in industrial organization. The problem considered is a monopolist choosing the characteristics for a new product to be launched into a market where consumer preferences change over time, according to a stochastic process, and are not directly observable.

The firm can perform market research but must process the results received in order to reach a decision. Using more information gives a better picture of the market at the time the market research was conducted. However, increasing the amount of information used also increases the time taken to make a decision, and market conditions change. Thus the firm chooses the optimal quantity of information and the hierarchy with which to process it jointly, by trading off the advantages of a better historical picture and a faster decision.

A simple model with a fixed price and a simple set of hierarchies is considered in chapter 3. The analysis is extended to a variable price model, with no constraint on the set of hierarchies, in chapter 4.

In the final paper, chapter 5, we focus on the labour market and examine the connection between hierarchies and wages. The paper first surveys the theoretical literature on supervision in hierarchies, known as the grand contract model. Although this literature is well known to those who work on hierarchies, it is not so well known in the labour literature, despite its relevance to the new jobs-based approach of Lazear(1995) and Baker et al(1995). The survey suggests a number of conjectures deserving empirical testing. This empirical analysis is performed on a unique Australian data set which includes human capital variables and variables on hierarchical position. The results indicate that both traditional human capital and newer institutional factors are important in wage determination. In particular, the new results show that height in a firm's hierarchy and the nature of the firm are both significant determinants of wages.

Chapter 6 concludes and indicates areas of future research.

1.3 Defining Hierarchies

The modeling of management structures in this thesis requires explicit specification of the relationships which connect people, for example, passing information to another manager or being monitored by a supervisor. Such structural relationships are described mathematically by graph theory. Luckily, only hierarchies need to be considered for the analysis in the following chapters, rather than more general graphs. A familiarity with definitions of various types of hierarchies is assumed in the main text. I provide here definitions for those readers unfamiliar with graph theory.

A graph $G = \{N, E\}$ consists of a set N of nodes and a set E of edges. The nodes represent the items of interest, in this case people, the edges represent the relationships between the nodes. An edge e is a pair such as (a, b) where $a, b \in N$. Note that there does not have to be an edge between all nodes.

To see how a graph can represent a simple relationship between people consider the following simple example. There are four people a, b, c and d; these will be the nodes of the graph. Let there be a relationship of trust between a and b, b and c, c and a and between b and d. The trust relationship can be represented graphically by drawing a line between any two people who trust each other. The resulting graph is shown in Figure 1.1.

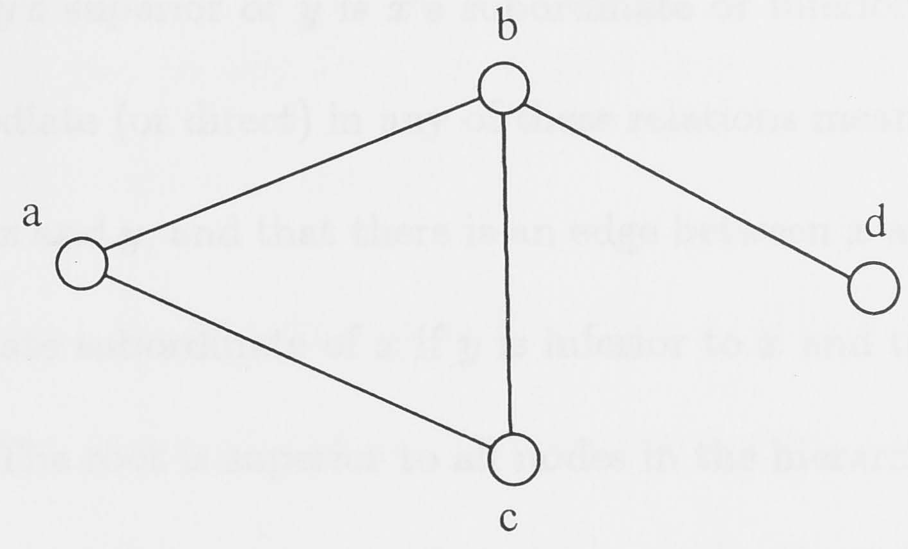


Figure 1.1: A simple graph

A graph is said to be connected if for all $y, z \in N$ there exists a series of edges such that $(y, x_1)\,(x_1, x_2)\dots(x_{n-1}, x_n)\,(x_n, z)$. A cycle in a graph is a series of edges $(y, x_1)\,(x_1, x_2)\dots(x_{n-1}, x_n)\,(x_n, y)$ such that each of the x_i's is distinct. An example of a cycle is $(a, b)\,(b, c)\,(c, a)$ in Figure 1.1. A connected graph with no cycles is called a tree. A hierarchy is a rooted tree. A rooted tree is just a tree with one particular node identified as the root. In the context of management structures the root of a hierarchy will be the boss or owner. Examples of a tree and a hierarchy are given in Figure 1.2.

For each node y, except the root r, in a hierarchy there is a unique series of edges, $(r, x_1)\,(x_1, x_2)\dots(x_{n-1}, x_n)\,(x_n, y)$, between y and the root. This fact generates two natural relations on a hierarchy, the inferior and superior relations. We say x is superior to y if x is a node in the series of edges that connects y to

x. y is inferior to x if and only if x is superior to y. If x is superior to y we will often say x is y's superior or y is x's subordinate or inferior. The inclusion of the word immediate (or direct) in any of these relations means that the relation holds between x and y, and that there is an edge between x and y. For example, y is an immediate subordinate of x if y is inferior to x and the edge (x, y) is in the hierarchy. The root is superior to all nodes in the hierarchy. The number of direct subordinates of a node x is called x's *span of control*, or, more simply, its *span*.

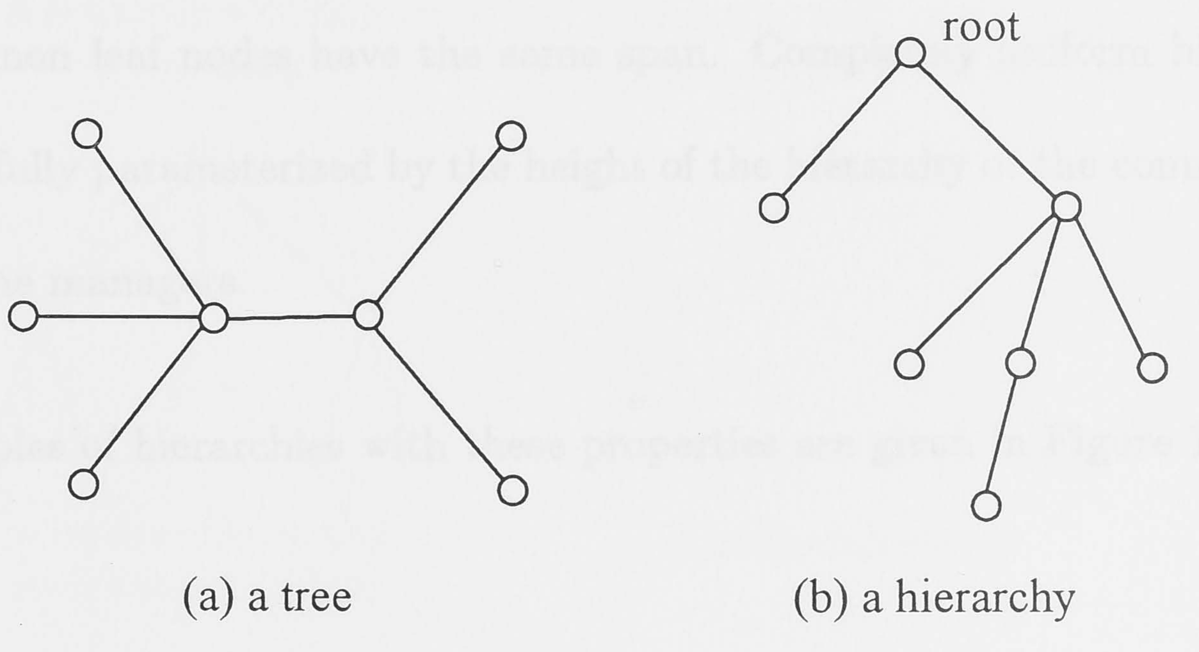


Figure 1.2: Examples of trees and hierarchies.

A *leaf* in a hierarchy is any node that has no inferiors. The *level* of a node x, is the maximum distance between x and any leaf inferior to x. Thus all leaves are at level 0. The *height* of a hierarchy is the level of the root.

The following regularity properties that a hierarchy can have are used elsewhere in the thesis:[3]

1. A hierarchy has no *skip-level reporting* if each immediate subordinate of any node at level l is in level $l-1$.

2. A hierarchy is *completely balanced* if it has no skip level reporting and if all the nodes in the same level have the same span. Completely balanced hierarchies are fully parameterized by the spans of the tiers and the height of the hierarchy.

3. A hierarchy is *completely uniform* if it has no skip level reporting and if all the non leaf nodes have the same span. Completely uniform hierarchies are fully parameterized by the height of the hierarchy or the common span of the managers.

Examples of hierarchies with these properties are given in Figure 1.3.

[3]These definitions are from Van Zandt(1996b)

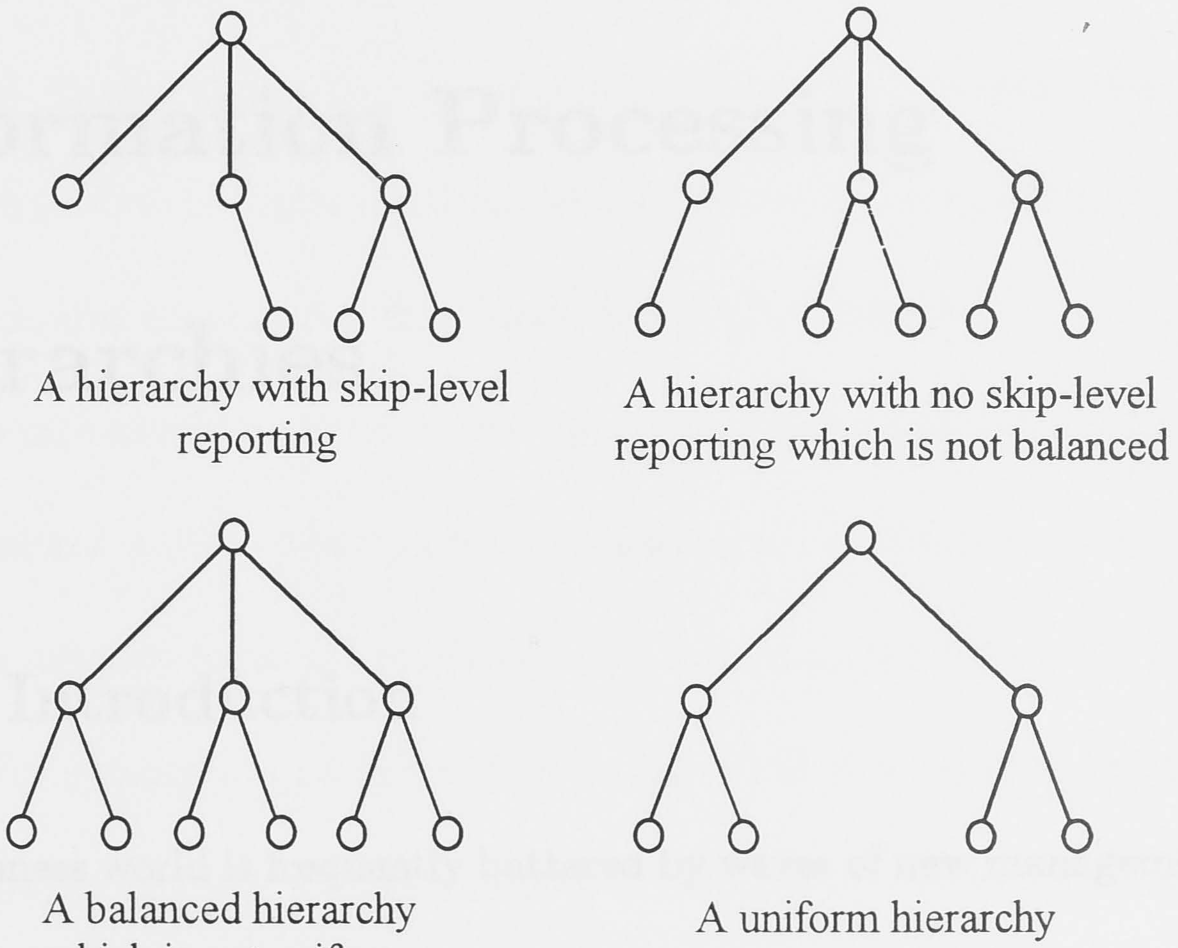


Figure 1.3: Regularity properties of hierarchies

Chapter 2

Information Processing Hierarchies

2.1 Introduction

The business world is frequently battered by waves of new management theories. Two very influential recent management theories have been re-engineering and contracting out. The motivation for these theories is that significant managerial inefficiencies can be overcome by reorganization. This raises the economically interesting question of what causes these inefficiencies to arise?

Inefficiency could arise solely as the result of employing more labor than is actually needed. Abstracting from incentive issues in this case, inefficiency would be eliminated by paying people only for the work needed (a piece-rate type regime) as opposed to all the time they are at work (a salary regime).

Costly waste labor would be eliminated and there would be no need to consider the shape of the management structure.

Re-engineering, however, is generally characterized by both downsizing and radical reorganization of the management structure. Implicitly at least, there is an assertion that gains exist from improving managerial structure as well. This is particularly clear in the debate over out-sourcing of government management activities. Beside the gains from only paying for required activities, there is also an assertion that the organizational structure will be more efficient, in some way, in a successful consultancy firm than in the government.

It seems natural to assume that being able to employ people for a few hours where needed would open up new organizational possibilities compared with the situation where people are employed at a fixed cost regardless of how much they work. For example consider a managerial bottleneck where on rare occasions one manager receives a large number of reports at the same time, but has an easy work load otherwise. It may not be worth hiring a new salaried employee to assist with this rare occurrence but moving to piece-rates could make such a reorganization of tasks profitable.

In order to consider these task assignment and hierarchical design issues, an economic framework is needed which is sufficiently general as to apply to a large number of situations, but will still correspond to some common, readily identifiable, management task. The information processing model pioneered in the work of Radner(1993), Radner and Van Zandt(1992) and Van Zandt(1995a) is used here as the basis for such a framework.[1] Working with information is a major area of employment in the economies of the developed world. Baumol et al.(1989), Gittleman and Wolfe(1995) and Radner(1992) all put the fraction of information or data workers in the U.S. economy at over 40% of those employed. Hence the information processing approach has wide application.

In this chapter, the employment side of the information processing model is extended to include piece-rate and salary regimes to address the following question: does changing the employment regime change the set of efficient management structures or just alter the cost? Surprisingly, it is found that for non routine decisions, the set of efficient management hierarchies is equivalent under differing employment regimes. Piece-rate regimes are superior in that they pay for fewer hours, not because they allow more efficient structures for dealing with information.

For example, consider a large manufacturing firm which periodically opens new plants. Should the firm have an in house department to decide on the type, location and size of each new plant or hire consultants for each project? It will be shown here that the set of efficient organizational structures is independent of the employment regime. Thus in choosing in house or out-sourcing the only differences the firm needs to consider are those that apply to incentives and

[1] The models of Bolton and Dewatripont(1994) and Keren and Levhari(1979, 1983, 1989) are closely related to the Radner-Van Zandt framework. See Van Zandt(1997b) for a discussion on the relationship between these models.

employment. Both alternatives will have the same set of potential management structures available to them.

Interestingly, changing the employment regime alters the returns to scale in information processing. In a piece-rate regime constant returns to scale are found for a linear loss function in contrast to decreasing returns to scale in Radner(1993).

The above results require analysis of the individual tasks people perform. This analysis is also used to extend information processing theory in a new direction. In the traditional information processing models people are completely reliable in the performance of an information processing task. That is, for a given set of inputs they will always produce the unique correct answer in a specified amount of time. This approach conflicts with other approaches to organizations such as the principal-agent model, where it is explicitly assumed that an agent's productivity is not directly observable. Two possibilities for the less than perfect performance of a task are considered: incorrectly processing information (error reliability) and being late in reporting the processed result (time reliability). In keeping with the current information processing literature, the affect of task assignment (shape of the hierarchy) on these reliabilities is considered. The relationship between reliability and task assignment is developed axiomatically from two well known management theory proverbs. Finally, this relationship is used to reduce the multiplicity of efficient hierarchies.

The two reliability concepts are quite natural and could readily be described as the result of either skill or effort, both of which suggest the consideration of incentives (wages are at present treated as a parameter).The concept of reliabilities provides a framework for future research into a synthesis of incentives and information processing.

2.2 Management Structures for the One Shot Mode

2.2.1 The Decision Making Process

Decision making can occur in many ways within an organization. It is generally based on information. Agents typically use two kinds of information in the decision making process: their own accumulated knowledge which has been learned from their experiences and new or previously unknown data. Information in the following analysis and discussion will refer to this second data type.[2]

This chapter focuses on how an organization takes new (and implicitly useful) information and uses it to make a decision. By making the organization the unit of analysis in the decision making process we are choosing to analyze decisions in which, potentially at least, a number of people can play a role. The number of people involved in a decision is endogenously determined given the quantity

[2]Although it is probably essential to have some prior knowledge of how the world works in order to be able to deal with new data, the accumulation and refinement of knowledge/skills is not the concern of this paper.

of information and the costs associated with decision making.

A number of people can be involved in a decision in two ways, either using the information collectively or separately. If they use it separately, then each person uses all the information to draw a conclusion. The final decision is then made by comparing the different conclusions, see Sah and Stiglitz(1986). There is no point in this arrangement unless different agents are going to draw different conclusions from the same information. This implies that either the process for using the information is non algorithmic in a way that precludes communicating the decision making procedure to another person, or there is some stochastic element to the algorithm. The first situation is observed in the real world when a group of people make a subjective decision such as which film to buy or which books to publish. The second situation is observed when people double check each others' work.

If the information is used collectively then the information is shared out and people work on separate parts of the decision, combining their partial conclusions into a decision. This aspect of decision making is commonly observed in organizations, where subordinates summarize data and write recommendations on aspects of a decision for the boss who in turn, weighs up the various reports in order to make the final decision.

This second, more common, type of decision making process is the one we shall consider here. In order to model this process we now make some of these decision making concepts more concrete.

We assume that the decision is a function of the information received. Thus the process of making the decision is analogous to the process of computing the value of the function. In order for the tasks in the decision making process to be spread amongst different people, it must be the case that the overall function is *separable* in some way. This is modeled here by assuming that the decision can be reached by adding in one extra piece of information at a time and that any splitting of the information into subsets to be dealt with by separate people, has no effect on the outcome. These can be expressed formally by the following assumptions.

Information: *Information is in discrete units, for example observations on stochastic processes such as weekly sales of different products by a retailer. A* **cohort** *consists of N units which arrive together to be processed to make one decision. The* **information set** *is all the information used, which may consist of many cohorts.*

Separability: *The decision is made by an associative binary operation.*

There is no reason to presume that these assumptions constrain this model to only simple, trivial calculations. For example choosing the maximum or minimum from a finite set is a binary associative operation, as is pattern matching. These are two common paradigms for describing decision making in an organization. Another example is estimating a distribution from a random sample. The separability assumption could be weakened to include other classes of decision functions. However there is no obvious direction in which to expand the description of decision functions since at present there is no underlying economic theory of the properties of decision functions.

The discussion at the start of this section indicates why separable decision problems are of interest. It does not indicate why we should ever expect to see the sharing of the decision making process amongst individuals. This occurs because processing information to make a decision takes time, and time is an important factor in the profitability of decision making. Sharing information between people allows them to work concurrently on different bits of the decision, reducing the total amount of time taken.

The assumption, that processing information takes time, is a very weak form of bounded rationality. Unlike the machine game literature (see Osborne and Rubinstein(1994, chapter 9)) it is not an assumption that people behave like automata. The only constraint placed on intelligence is that the information processing to be performed is sufficiently complex as to require a discernible amount of time. This amount is standardized to be one period of time per unit of information.

Although necessary for the analysis that follows, the assumption that information cannot be assimilated instantaneously is a plausible description of human decision making, as the following simple example shows. Consider an individual reading reports on investment opportunities in six different countries, in order to choose one in which to invest. For this individual, like everyone else, it will take a non zero amount of time to read each report.

If we assume that it takes one unit of time, say an hour, to read each report then we can reduce the total amount of time taken to make the decision by splitting the task between two people. In the original situation it took one person six hours to read all six reports and to decide which country is most suitable. If the two people start work together and each one reads three reports then it takes three hours to find the best country in each subset. The second person then hands the report for the best country in the second subset to the first person. The first person then takes another hour to read this report and to decide if this country is better than the best country from their subset of reports. The best country is then found in four hours instead of six. Sharing the processing so that work can be done concurrently reduces the time taken to make a decision. If we think of the reports being about movements in the exchange rates for a set of currencies, then the importance of making a quick decision becomes obvious.

The time element is captured explicitly here by assuming that individuals in the organization are modeled as *processors*. A processor has an *in-box* and a *register* (which together comprise its memory). Let $f(x, y)$ be the binary associative function being used for this decision and let y be the contents of the processors register. Then in one *period* of time a processor can take one piece of information, x, from its in-box, perform $f(x, y)$ and store the result in its register. It can then write the contents of its register (called a *report*) to the in-box of any processor to which it is connected. A *decision* is made by the hierarchy when the contents of a processor's register are outputed rather then being written to another processor. Together processors and connections form a network. A program specifies when each processor writes to another and when a decision is made. Initially all registers are set to zero and the information is assigned to some of the in-boxes.

In the following, only networks which correctly process all the information assigned to them in finite time will be considered.

A processor is only a partial description of a person, only those features pertinent to information processing are considered. This is analogous to the situation in principal-agent theory where only those aspects of the agent relevant to incentive considerations are modeled. Thus just as a principal-agent model doesn't indicate how to organize a firm to produce decisions quickly, the information processing model doesn't indicate what the employment contract should be. A complete theory of organization needs to integrate at least these two approaches.

It might seem that these processors are much less intelligent than real people, this is not the case. As discussed above, taking time to deal with information is certainly a human characteristic, in fact the normalization of only one calculation per period does not preclude processors actually being faster than people. In order to actually construct networks to make decisions, the binary associative operation must be algorithmic, see Meagher(1996). That extra requirement is not binding here since the actual method a processor uses to perform the functional evaluation of f is not considered here. This leaves only the algorithmic way in which information is transferred and new information is "read in". The original model of Radner and Van Zandt can be extended to include communication costs (see Bolton and Dewatripont(1995)), this extension is not considered here. As always the appropriateness of the assumptions made here is a matter for empirical testing.

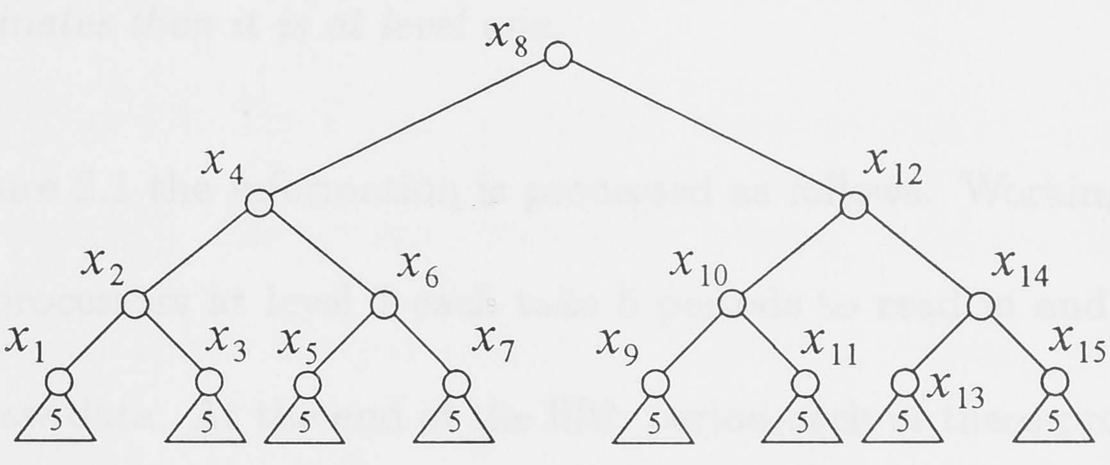


Figure 2.1: A 3 level hierachy using 11 processors on 40 pieces of information with a delay of 11 cycles.

Figure 2.1 shows a hierarchy with 11 processors (represented by the circles). The lines between the processors indicate connections which are used to pass information. A triangle indicates a group of raw data that a processor is dealing with. The number in the triangle indicates the number of pieces of information, in this case 5. For the purpose of this example we shall assume that the information set contains only one cohort (this is known as the *one shot mode)*.

We will be concerned primarily with hierarchies. A network is a hierarchy if it contains no *circuits (cycles)*. A circuit/cycle is a series of connections which, starting at a processor, lead back to that processor. For convenience and precision the following definition of levels in a hierarchy will be used throughout the chapter.

Definition 1 *If a processor has subordinates then the level of the processor is one greater than the maximum level amongst its subordinates. If a processor has no subordinates then it is at level one.*

In Figure 2.1 the information is processed as follows. Working at the same time the processors at level 1 each take 5 periods to read in and process their assigned raw data. At the end of the fifth period each of these processors sends the result of its calculation to its superior at level 2. Hence each processor at level 2 processes 2 pieces of information which takes another 2 periods. Thus at the end of the seventh period each level 2 processor passes its output to its supervisor at level 3. Similarly for level 3 and 4, except at the end of the eleventh period the processor at the top of the hierarchy does not pass on the results of a calculation, but rather produces a decision based on all the information it has received. Hence this hierarchy takes 11 periods to produce a decision based on 40 pieces of information using 15 people.

Processor activity can be represented diagrammatically with an activity dia-

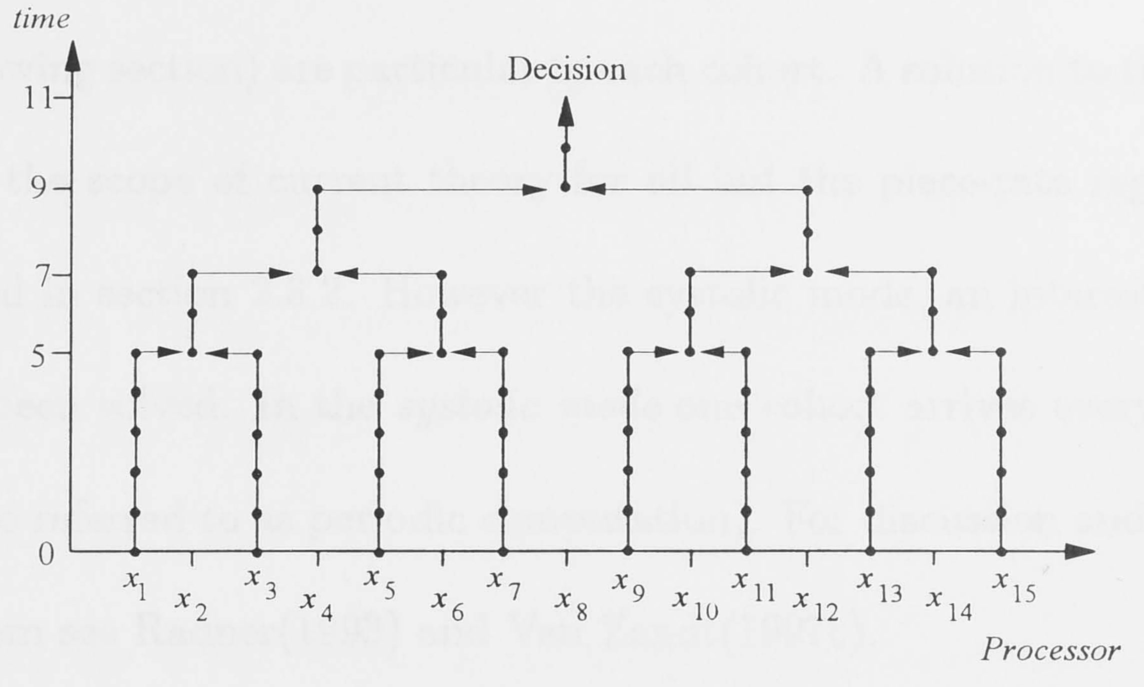


Figure 2.2: Activity diagram for the binary hierarchy of Figure 1.

gram, see Figure 2.2. There is a line for each period of time in which a processor is active (working). The arrows indicate the passing of information by processors. Generally information is passed to another processor, however for the processor at the top of the hierarchy this represents making a decision.

2.2.2 The Information Set

The example from the previous section demonstrated how an information processing hierarchy made a decision on one cohort of information. This is called the one shot mode.

Multiple cohorts of data are also of interest. Without relaxing the assumption that one decision is to be made based on each cohort, multiple cohorts gives rise to a complex environment. The most general case is that ν cohorts arrive in any one period (ν stochastic) and that the costs associated with processing (defined in the following section) are particular to each cohort. A solution to this problem is beyond the scope of current theory for all but the piece-rate regime, which is discussed in section 2.3.2. However the systolic mode, an interesting special case, has been solved. In the *systolic mode* one cohort arrives every T periods (this is also referred to as periodic computation). For discussion and solution of this problem see Radner(1993) and Van Zandt(1997c).

The goal of this chapter is to enrich the description of processors in these models by making their payment more economically meaningful and by allowing them to be stochastic in the performance of their tasks. Given this goal the one shot mode is considered because the exposition of the definitions and theory in sections 2.3 and 2.5 is much clearer.

2.3 Efficiency of Management Structures

In order to determine the efficiency of a management structure there must be measures of the cost and performance of the structure. Section 2.3.1 defines these measures, shows how they can be calculated and gives a definition of efficiency based upon them. Section 2.3.2 examines the relationships between efficient hierarchies under different employment regimes.

2.3.1 Costs in Decision Making

The programmed network model allows the calculation of the relationship between the number of people, the time taken to make a decision (referred to as delay), and the amount of information used in making the decision. From this information a production set can be generated. For details of this see Radner(1993).

Applying the profit maximization methodology, the optimal management structure for a certain situation would be the one giving the highest expected profit. However to determine profitability, the management structures need to be operating in a specific market environment, for example see Meagher(1996). A more general approach is to use a loss function to describe the costs associated with a management structure and the decision it makes. Interest has centered around the trade off between the number of people P and the delay C. Hence the use of the linear loss function of Radner(1993) shown in equation 2.1

$$L = \lambda C + \phi P. \tag{2.1}$$

Loss, L, for a fixed N comprises of a total labor cost ϕP and a cost due to delay λC. The appropriateness, or otherwise, of the linearity of the delay cost depends on the decision problem and some alternative formulations are discussed in Radner(1993).

The contribution of this chapter is not to analyze further the cost of delay, but rather to reformulate the labor cost term in a more economically meaningful way. As it stands the cost of labor is a fixed cost per person, independent of the amount of time each person is employed.

It could be argued that ϕ is just a lump sum payment, equivalent to some effective per period wage w_e multiplied by C the number of periods it takes to make the decision. However C varies between different management structures, so that w_e would have to vary inversely with C in order to keep ϕ constant.

Fortunately, the detailed construction of this model provides all the time related information needed to build a realistic labor cost term. The programmed network model does not imply a labor market, hence employment regimes (the number of periods a person gets paid for) are taken as exogenous.

The measure of the labor input (which is analogous to hours), H, for a given network depends on the employment regime (expressed by the function $H(regime)$). We consider the following three regimes:

- The *salary* regime, where each of the P processors is employed for the duration of information processing. In the one shot mode that is C periods, hence $H(salary) = CP$, (where C is the delay in periods between the start of the information processing and the output of a decision).

- The *processor* regime of Radner(1993), where $H(processor) = P$.

- The *contract/piece-rate* regime, where processors are only employed for the periods in which they are active. I show in (Lemma 6) that for an efficient one shot network under piece-rates $H(piece - rate) = N + P - 1$.

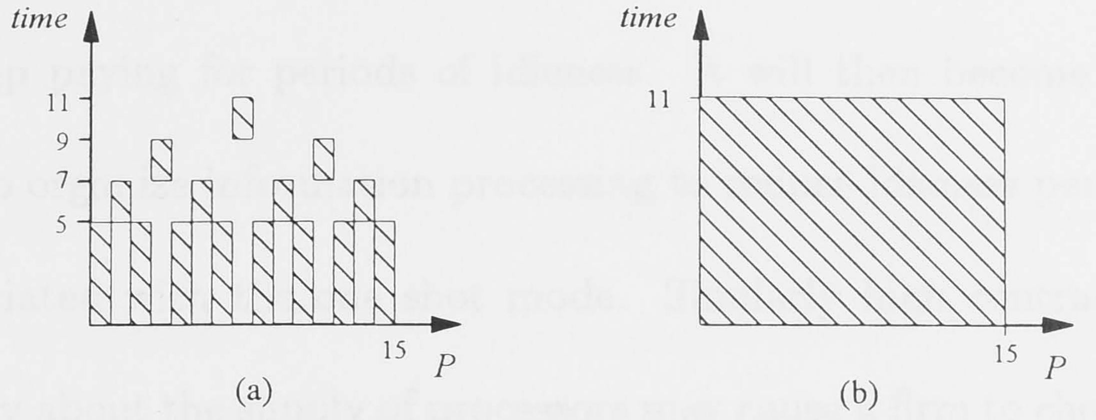


Figure 2.3: The labour input for the hierarchy in Figure 1. can be represented as areas on the activity diagram. (a) The contract/piece-rate regime. (b) The salary regime.

The contract/piece-rate and salary regimes have intuitive graphical interpretations in the activity diagram. Consider again the example hierarchy of Figure 2.1 and its activity diagram, shown in Figure 2.2. Under the contract/piece-rate regime the firm pays for the total number of active periods, that is the total number of periods of work done. The number of work periods is found by changing each line on the activity diagram to a bar of width one and then summing the areas. This is applied to the binary hierarchy example in Figure 2.3(a), where the total area, and hence the total labor input, is 54.

Under the salary regime the firm has to pay each individual for the entire duration of the process, which gives a labor input of CP for the one shot mode. This comes to $11 \times 15 = 165$, which is the shaded area shown in Figure 2.3(b). The salary regime is most intuitively plausible under multiple cohort situations. Consider for example the systolic mode. Institutional reasons may preclude hiring people to work intermittently under the piece rate regime, thus the firm will end up paying for periods of idleness. It will then become important to attempt to organize information processing to reduce idleness per cohort below that associated with the one shot mode. Similarly high contracting costs or uncertainty about the supply of processors may cause a firm to choose to employ people even when they are idle between cohorts.

Clearly there can be significant differences in the quantity of labor employed under the different regimes. The importance of considering more than just the number of processors in describing the labor input was independently observed in Van Zandt(1997c) and Meagher(1996).[3]

Before any results are derived a few more definitions are needed.

For a particular regime let $w(\mathit{regime})$ denote the wage or payment to each unit of labor (as measured by $H(\mathit{regime})$).

Also assume that the loss function $L(H(\mathit{regime}), C)$ is linear

$$L = w(\mathit{regime})H(\mathit{regime}) + \lambda C \tag{2.2}$$

where λ is the loss/cost per extra period of delay.

[3]This paper focuses on the one shot mode, the solution of the multiple cohort problem under piece-rates and stochastic performance (reliabilities) to reduce the multiplicity of equilibria in these situations. Van Zandt(1997c) develops this framework in a complimentary, direction, extending the analysis of fully reliable processors by applying a salary type regime to periodic computation. In order to model networks that perform periodic computation Van Zandt(1997c) is much more detailed in his description of communication, leading to a more complex but still deterministic description of a processor.

Definition 2 *The* **quartet** $\langle N, H(\cdot), C, \text{regime}\rangle$ *denotes the performance of a network.*

Definition 3 *A network ($\langle N, H(\cdot), C, \text{regime}\rangle$ quartet) is* **efficient** *if for a given quantity of information N, and a given employment regime, it is not possible to decrease $H(\cdot)$ without increasing C, or vice versa.*

Radner(1993) includes a formula which gives the minimum C (denoted $\widehat{C}$) and a method for finding the network (the method in fact gives a hierarchy) with that $\widehat{C}$, for a given N and P. The formula is

$$\widehat{C} = \left\lfloor \frac{N}{P} \right\rfloor + \lceil \log_2(P + N \bmod P) \rceil \,. \tag{2.3}$$

2.3.2 Equivalence of Efficient Sets

The following theorem establishes the surprising result that the sets of efficient quartets under a number of employment regimes are unique. This contradicts the intuition that organizations would be more efficient if people could be employed for odd hours, where needed, instead of in large continuous blocks.

Theorem 4 *For the one shot mode:*

(a) The quartet $\langle N, P, C, \text{processor}\rangle$ is efficient if and only if $\langle N, N+P-1, C, \text{piece}-\text{rate}\rangle$ is efficient.

(b) If $\langle N, PC, C, \text{salary}\rangle$ is efficient then $\langle N, P, C, \text{processor}\rangle$ is efficient.

(c) If $\langle N, P, C, \text{processor}\rangle$ *is efficient and there does not exist* C' *and* P' *such that* $PC = P'C'$, $C' < C$ *and* $\langle N, P', C', \text{processor}\rangle$ *is efficient, then* $\langle N, PC, C, \text{salary}\rangle$ *is efficient.*

This theorem is proved at the end of the current section.

Van Zandt(1995a) also considers the relationship between efficiency under the processor regime and under the piece-rate regime for the one shot mode (although not the salary regime). Although related, the statement of the result and the proofs are different. The fundamental difference in approach is that here the results are based directly on the performance of the networks (the quartets), while in Van Zandt(1995a) the execution of tasks and the details of communication are emphasized, (leading to more detailed and involved proofs). It is not possible to be more precise without restating the Van Zandt model. Intuitively the Van Zandt framework necessitates the explicit consideration of networks which contain the redundant passing of messages. These are efficient under the processor regime but not under the piece-rate regime. These extra connections can also allow non hierarchical networks to be efficient under the processor regime. These problems are neatly circumvented here by considering the relationships between the efficient quartets under the different regimes rather than considering specific hierarchies. Corollary 5 then relates the results efficient quartets back to hierarchies that achieve them.

Corollary 5 *An efficient quartet, under any of the employment regimes, can be achieved by a hierarchy.*

Proof.

It follows from the construction of hierarchies in the proofs of the three parts of Theorem 4 that the one shot efficient hierarchies of Radner(1993) are efficient under all three regimes. ■

Theorem 4 says that regardless of the regime under which people are employed, the set of efficient ways in which to organize them in the one shot mode is in general the same.[4] That is, there can be no gains in efficiency from being able to employ someone for just a couple of periods as opposed to having to employ them at the same cost as everyone else. The piece-rate regime hierarchies will always pay for less periods of work than the salary regime hierarchies but this is the only gain - the organizational technology is the same.

The proof of Theorem 4 is rather involved and is left for the interested reader until the end of this section. The intuition however comes from the following lemma.

Lemma 6 *If all processors except the top processor write exactly one subtotal to another processor then $H(piece - rate) = N + P - 1$ in that network, which is a hierarchy.*

Proof. Lemma 6.

[4]I conjecture that the extra condition in Theorem 4(c) (on C' and P') rarely, if ever holds.

Each of the N pieces of information goes to only one in-box. Thus each piece of information is read in only once. This takes one active period per piece of information, that is N periods. By assumption each of the $P-1$ processors below the top processor writes a subtotal once to another processor. It then takes one active period to read in each of these $P-1$ subtotals. Hence we have $H(piece-rate) = N + P - 1$. Since an output is produced by the P processors they must all be connected into the network by the $P-1$ connections over which subtotals are passed. Hence there are only enough communication connections to form a hierarchy, and not enough to form any cycles. ■

For all regimes in the one shot mode, every processor receives raw data (Lemma 7). Hence every processor has a report to send. Lemma 6 provides a lower bound on the active cycles required to do this, regardless of the regime. Using this it can be shown that there cannot be any efficiency gains between regimes because this common technology underlies all the regimes.

Lemma 7 *Every processor in an efficient* $\langle N, H(.), C, \text{piece}-\text{rate}\rangle$ *hierarchy receives raw data.*

The proof of this lemma is contained in the appendix.

Theorem 8 *An efficient solution to any multiple cohort problem under the piece rate regime is to assign a one shot efficient hierarchy to each cohort.*

Proof.

By assumption there need be no interactions between processors working on separate cohorts. Under piece rates there are no idle periods between processing cohorts to minimize and hence the processing of each cohort can be considered separately. Thus efficiency is obtained overall if each cohort, considered in isolation is processed efficiently. This is achieved by using a one shot efficient hierarchy to process each cohort.[5] ■

The assumptions of associativity and one decision per cohort imply that there are no inter-relations between the processing of cohorts. That is, there is no need to worry about one decision being made before some other decision or about who is assigned to each cohort. The piece-rate regime leads to separability in labor cost. The labor cost of processing a cohort is just the number of active cycles needed to process it. There is no effect from when any other cohort arrives or how big it is, on the labor cost of processing a particular cohort.

In fact there is no need for the cohorts to have the same w or λ, or for the cohorts to be of the same size, or for the number arriving in any period to be known in advance since the choice of one shot efficient hierarchy for each cohort is independent of the choices for all other cohorts. These very strong results will not hold in general if there are planning costs, or if computation needs to

[5]The special case of this theorem, applying to the systolic mode, was pointed out to me during a conversation with Roy Radner and Tim Van Zandt in the spring of 1995 and is noted in Van Zandt(1996,p37). The much more general multiple cohort case is particular to this paper.

be performed by one network, or if people have to be paid under some sort of salary regime. In these circumstances the decision problems will need to be solved jointly since there would be interactions.

The proofs of the main results follow.

Proof. Sufficiency of Theorem 4(a).

Radner(1993) shows that for each efficient $\langle N, H(.), C, processor\rangle$ quartet there exists a hierarchy, which achieves that quartet, in which each processor except the top processor writes exactly one subtotal. By Lemma 6 these efficient $\langle N, H(.), C, processor\rangle$ hierarchies use $N+P-1$ active periods. We now consider a piece-rate regime hierarchy, using $N+P-1$ active periods to process N pieces of information. Any such hierarchy must use N periods to read in the raw data hence it can use at most $P-1$ periods for adding up subtotals. The top processor reads in raw data and hence has its own subtotal, to which all other subtotals must eventually be added. Hence there can be at most P subtotals. By Lemma 7 each processor receives raw information and hence has its own subtotal. Thus there can be at most P processors.

It remains to show that an efficient $\langle N, N+P-1, C^*, piece-rate\rangle$ hierarchy does indeed use P processors, but has the same delay as the associated efficient $\langle N, H(.), C, processor\rangle$ hierarchy.

If it were possible to subdivide and add up the information using less than $N+P-1$ active periods to give a delay of $C^* \leq C$ then the $\langle N, H(.), C, processor\rangle$ hierarchy could not have been efficient (since reducing the number of active periods means reducing the maximum number of processors used below P). This would imply that there is some way to organize less than P processors to give a delay of at most C, contradicting the assumed efficiency of the $\langle N, P, C, processor\rangle$ hierarchy.

If it were possible to use P processors in a $\langle N, N+P-1, C^*, piece-rate\rangle$ hierarchy to give a delay of $C^* < C$, then it would certainly be possible to do this with P processors under the processor regime because there would be no limit on the number of active periods used. Hence $C^* = C$. ■

Proof. Necessity of Theorem 4(a)

$N+P-1$ active periods are used in an efficient $\langle N, H(.), C, piece-rate\rangle$ hierarchy. Rearranging, $P = H - N + 1$. Using this P and the same N we can find a $\langle N, P, C^*, processor\rangle$ hierarchy, where C^* is the minimum delay possible given N and P.

By assumption, the $\langle N, N+P-1, C, piece-rate\rangle$ hierarchy was efficient so $C \leq C^*$. Otherwise we could construct a hierarchy from the $\langle N, P, C^*, processor\rangle$ hierarchy, using $N+P-1$ active periods but having shorter delay.

Also $C \geq C^*$otherwise we could construct a hierarchy from the $\langle N, N+P-1, C, piece-rate\rangle$ hierarchy which would use P processors but having a shorter delay than C^*, violating the choice of $\langle N, P, C^*, processor\rangle$ as a hierarchy with minimum delay for the given N and P.

Since $C = C^*$ it follows that there is no $\langle N, P^*, C, processor\rangle$ hierarchy where $P^* < P$, because such a hierarchy would use less than the $N+P-1$ active periods to give a delay of C violating the assumption of $\langle N, N+P-1, C, piece-rate\rangle$ as an efficient hierarchy. ■

Proof. Theorem 4(b)

Dividing the labor cost $H(salary) = PC$ by C gives the number of processors used. The proof then proceeds as in the proof of Theorem 4(a), to show that by efficiency, P and C are the same under both measures of labor input. ■

Proof. Theorem 4(c)

By assumption there is no other processor efficient hierarchy with the same labor input under the salary regime as $\langle N, P, C, processor\rangle$ which has a lower delay. Applying arguments similar to those used in the proof of part (a) of this theorem, there cannot be a salary efficient hierarchy having both a lower input and a lower delay than $\langle N, P, C, processor\rangle$, under the salary regime. If such a hierarchy did exist we could use it to construct an efficient processor hierarchy using less processors and incurring less delay than $\langle N, P, C, processor\rangle$. This would violate the assumption that $\langle N, P, C, processor\rangle$ is an efficient processor hierarchy. Thus under the salary regime $\langle N, P, C, processor\rangle$ is an efficient hierarchy. ■

Having examined the efficiency of one shot hierarchies under the various regimes we now proceed to consider the choice of hierarchy.

2.4 Efficient Hierarchies and the Efficient Frontier

We now examine the characteristics of efficient hierarchies under the differing employment regimes. Section 4.1 makes comparisons between the efficient frontiers in a manner similar to the standard isoquant-isocost framework. Section 4.2 uses the new results on hours to see how hours worked grows relative to numbers of people employed as the quantity of information grows, and hence to examine the returns to scale.

2.4.1 Non Convexity of the Efficient Set.

Figure 2.4 shows the efficient frontiers for the salary and piece rate regimes for the case $N = 40$. This shows the standard relationship between the two frontiers. They touch when the hierarchy has one processor and the piece rate regime frontier is below the salary regime in all other cases. Thus given the same effective per period wage and cost of delay, the piece rate regime with the same number of employees is always no more expensive than the corresponding salary regime, and the piece rate regime will be strictly cheaper for sufficiently many employees.

Comparing the preferred efficient piece rate regime hierarchy to the preferred salary regime hierarchy (assuming effective per period cost are the same) we observe more people are employed, with a lower average total income in a faster acting organization than under the salary regime.

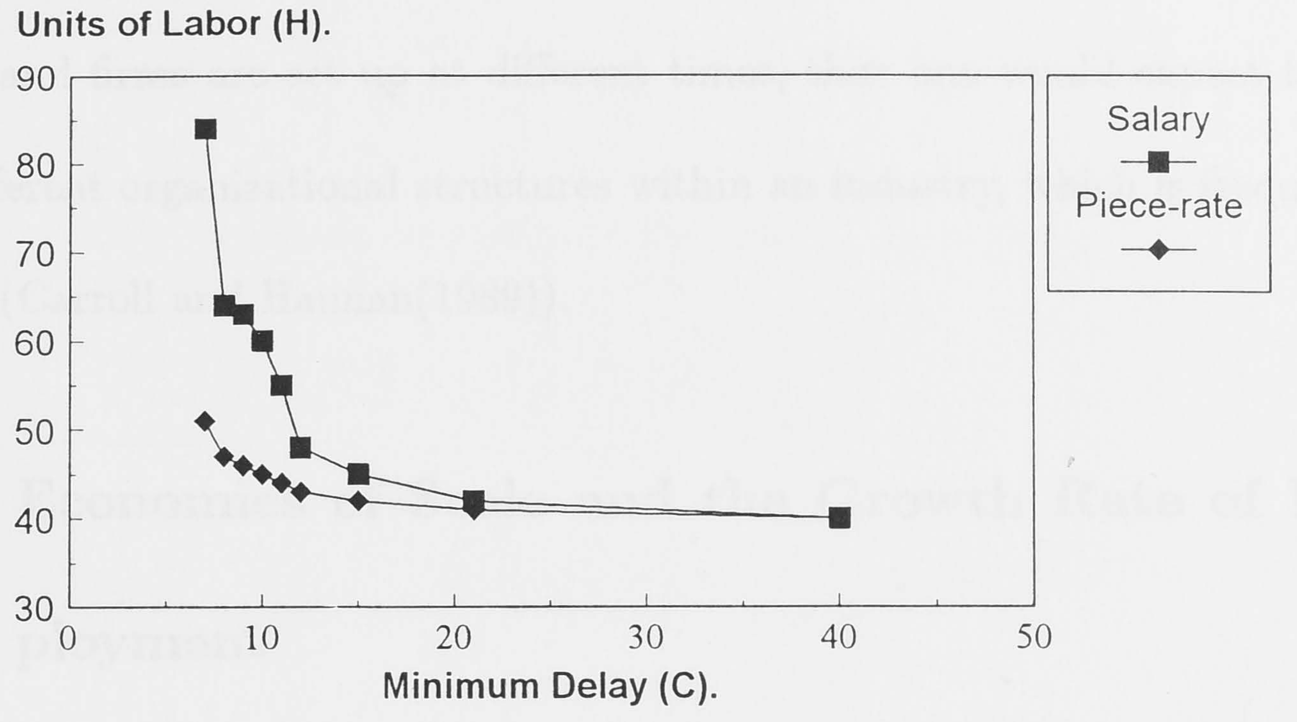


Figure 2.4: Efficient frontiers for salary and piece-rate regimes

Since the efficient frontier of performance quartets consists of only points it is possible that more than one quartet minimizes a given loss function. In which case one of the refinements from Section 5 can be applied to select a unique point.

The underlying non convexity of the relationship between P and N for efficient hierarchies that Radner(1993) observes carries over to the piece-rate and salary regimes. In fact the salary regime is very non convex in some regions. The non convexity occurs near quartets with the smallest delay. Technically this implies that in this region small changes in relative prices could lead to large shifts around the frontier, with corresponding major restructuring in the hierarchy and changes in employment. This would be difficult to test directly since there are generally other costs to reorganization which produce inertia (Carroll and Hannan(1989)). However, given inertia, if relative prices (λ and w) are variable and firms are set up at different times, then one would expect to see quite different organizational structures within an industry, which is frequently the case (Carroll and Hannan(1989)).

2.4.2 Economies of Scale and the Growth Rate of Employment

I now examine how H and C depend on N for efficient hierarchies under the salary and piece-rate regimes. This builds on the earlier work of Radner(1993) which analyses how H and C depend on N for the processor regime. These results can be interpreted as returns to scale in information processing, where N is the scale of the information processing.

From Radner(1993) if N, P and $\frac{N}{P}$ are large then the following approximation holds for $\widehat{C}$ (minimum delay).

$$\widehat{C} \approx \frac{N}{P} + \log_2 P \tag{2.4}$$

The loss function (total cost) for the piece-rate regime is

$$L = w(N + P - 1) + \lambda C \tag{2.5}$$

Using the approximation in Equation 2.4 gives

$$L = w(N+P-1) + \lambda(\frac{N}{P} + \log_2 P) \tag{2.6}$$

$$= w(N+P-1) + \lambda(\frac{N}{P} + \frac{\ln P}{\ln 2}) \tag{2.7}$$

The first order condition to minimize the loss, L, for a given N is

$$w + \lambda\left(-\frac{N}{P^2} + \frac{1}{P\ln 2}\right) = 0 \tag{2.8}$$

or equivalently

$$\alpha P^2 + \frac{P}{\ln 2} - N = 0, \qquad \alpha \equiv \frac{w}{\lambda}. \tag{2.9}$$

Solving this quadratic for the positive root gives

$$P = \frac{-\frac{1}{\ln 2} + \sqrt{\left(\frac{1}{(\ln 2)^2} + 4\alpha N\right)}}{2\alpha} \tag{2.10}$$

Thus as N increases without bound it follows

$$P \sim \left(\frac{N}{\alpha}\right)^{1/2}, \qquad C \sim (\alpha N)^{1/2}, \qquad H \sim N, \qquad \min L \sim wN. \tag{2.11}$$

These results indicate P and C growing at the same rate asymptotically as in the processor regime, see Radner(1993). However by applying the methodology of hours rather than heads to measure the labor input, we see that under the piece-rate regime labor is actually increasing at the rate of N, rather than $\sqrt{N}$, as found in Radner(1993). As a consequence total cost also grows at the higher rate of N, which means asymptotically constant returns to scale.

The asymptotic behavior of the average number of periods of activity per person is

$$\frac{H}{P} \sim N \times \left(\frac{\alpha}{N}\right)^{1/2} = (\alpha N)^{1/2}. \tag{2.12}$$

This leads to the empirical implication that in an information processing team, working on one project at a time, the average hours worked per person is of the order of the square root of the amount of information being considered. One possible real world situation in which to test this is large legal cases. In such cases a team sets to work on sifting through evidence in order to compile a case. The number of hours and the number of people would be recorded for billing purposes and the quantity of information could be approximated by the number of pages of evidence considered.

2.5 Uniqueness of Equilibria

There can fail to be a unique equilibrium hierarchy for two reasons. First non convexities in the efficient frontier can give rise to two or more performance quartets that minimize the loss function. The second reason for the non uniqueness of equilibria is that the efficient performance quartet may correspond to more than one hierarchy.[6] The points on the efficient frontier correspond in general to sets of hierarchies rather than individual hierarchies. For example, Figure 2.5 shows the four efficient hierarchies using seven processors (for $N = 7x$, $x \in \mathbf{Z}$ and x sufficiently large).[7] .

[6]To avoid confusion we focus on piece-rate efficiency in this section.

[7]There are two parts to processing information: preprocessing and overhead processing. During preprocessing each processor deals with the raw information it has been assigned. Overhead processing is the bringing together of information from the P processors. Since in-

The existence of multiple efficient hierarchies does not invalidate any of the previous results since uniqueness was not implicit in any of the proofs. However, an actual firm must choose a specific hierarchy rather than just a performance quartet. Although this could be done randomly, it is more interesting and informative to consider refinements to the set of equilibria that reflect other aspects of the design of hierarchies. The four equilibrium refinements which I present here are particularly interesting because they extend the analysis to situations in which the performance of a processor is no longer completely reliable.

There are many extensions to this model which would capture pertinent features in hierarchy design, such as standardization of tasks, tournament for pay setting/promotions and learning by doing. Instead of these we focus here on processors whose performance is stochastic. In particular people can either make mistakes in the reports they pass or can be late in passing their reports.

Firstly, I will consider processors that can make mistakes in their information processing tasks.

2.5.1 Uniform Error Reliability.

Assume that for each processor there exists an identical probability θ that the processor passes a report without an error (called the *reliability*), and that the

formation is assigned evenly for preprocessing, see Radner(1993), the overhead processing can be considered independent of N, for sufficiently large N. The problem for overhead processing is how to bring information from P sources together as quickly as possible.

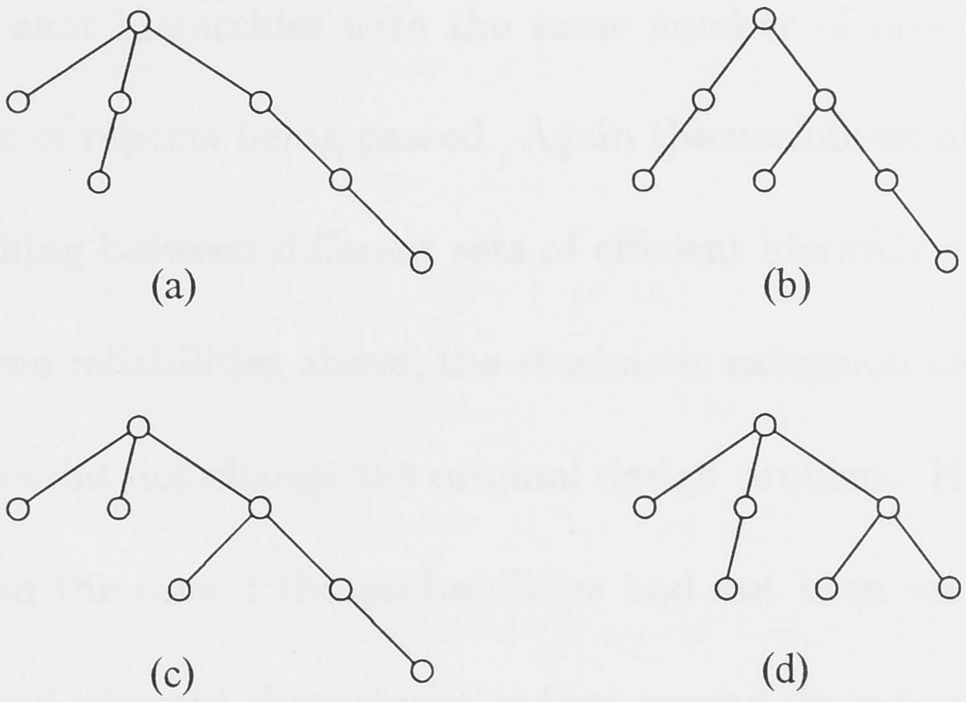


Figure 2.5:

realization for each processor (whether it makes a mistake or not) is independent.

Then for each of the efficient hierarchies in Figure 2.5 the probability that they compute the correct decision is θ^7. Hence this refinement fails to differentiate between equilibrium hierarchies with the same number of processors. This refinement is only useful when there is more than one efficient performance quartet, in which case it selects the quartet with the smallest number of processors. This is a formalization of the common idea in management theory that smaller and hence simpler is better: " ..things need to be kept simple if the unit is truly to pull together." Peters and Waterman(1982,p306)

A similar type of reliability is to attach the probabilities to the reliability of the messages (a message may become garbled). In this case the intuition is to minimize the number of times information is passed from one processor to another. Again this has no impact on a single set of efficient hierarchies, because efficient one shot hierarchies with the same number of processors all have the same number of reports being passed. Again the usefulness of this refinement is in distinguishing between different sets of efficient hierarchies.

For the two reliabilities above, the stochastic extension to the description of the processors did not change the original design problem. However that would not have been the case if the probabilities had not been identical. If different individuals had inherent characteristics that caused them to perform differently, regardless of their position, then task assignment to individuals would be important and would need to be considered in design (an example of this in a different kind of hierarchy model is Rosen(1982)). This is a much more complicated problem because the distribution of characteristics has to be known before design, or the organizational design has to include elements for determining an individuals characteristics.

The remainder of this chapter will not focus on differences in performance which spring from an individual's personal attributes. Instead the refinements concentrate on how the structure in which one operates might affect one's performance. The two aspects of performance to be considered are reliability and potential for causing delay.

2.5.2 Positional Error Reliability

Consider now the case in which position in the hierarchy, and hence the tasks one is assigned, affect one's reliability. Let the probability that a processor correctly processes its inputs, $\theta(I) \in (0, 1]$, be a function of I the number of information processing tasks the processor performs. $\theta(I)$ is called a processor's *primary reliability*. One natural intuition about $\theta(I)$ is that a person is more likely to make a mistake if they have more things to do (once they have learned their task). This intuition is modeled by assuming that $\theta(I)$ *is decreasing in* I. Also assume that the probability of errors in processing are *independent* for different processors.

A processor can also produce an erroneous report because it received incorrect reports. Thus the concepts of the *overall reliability* of a processor is also needed. Let R, the overall reliability of a processor, be the probability that the processors report is correct.

Under these assumptions different hierarchies will in general have different overall reliabilities, even if they have the same N and P. Thus the overall reliability of a hierarchy (and hence of the decision it makes) should also be treated as a choice variable. However the loss function will be sensitive to the specific functional form of $\theta(I)$, and existing information processing theory gives no indication of what specific form $\theta(I)$ should take.

In order to leave the loss function unchanged, we consider sequences of primary reliability probabilities which tend to full reliability ($\theta_n(I) = 1$) in the limit ($n \to \infty$) but fall in the interval $(0, 1)$ otherwise. This is very similar to the trembling hand refinement in Game theory. This assumption makes the choice of equilibrium hierarchy a two step process. First choose an efficient quartet, then choose the most reliable hierarchy in this quartet. Besides being tractable this method is also attractive because it requires a weaker assumption about $\theta_n(I)$ - it will now only be necessary to rank the reliabilities of efficient hierarchies with the same P. We now consider some economically meaningful ways in which to rank the reliabilities and the corresponding assumptions about the curvature of $\theta_n(I)$.

For a given N and P, the total number of information processing tasks across all processors in an efficient hierarchy is fixed at $N+P-1$ (see Lemma 6). Thus for a given N and P the maximum reliability of a one shot efficient hierarchy is determined by the allocation of the $N+P-1$ information processing tasks across the P processors. In order to rank the reliabilities of a set of hierarchies, further assumptions about $\theta_n(I)$ are necessary. Strict log concavity and strict log convexity are of particular importance as Theorem 9 shows.

Before stating the theorem the two new concepts of *span dominance* and *chain dominance* need to be introduced.

As Simon(1953) observes, organizational theorists have certain proverbs about how to best structure hierarchies. In particular, large spans of control are bad and long chains of command are bad. Simon calls these proverbs because they are opposite and unquantified statements. For a fixed number of people small spans of control necessarily imply long chains of command and vice versa. Obviously a good organizational design has to find a trade off between the two, the proverbs offer no assistance in determining this trade-off. These two proverbs are formalized in the following two definitions.

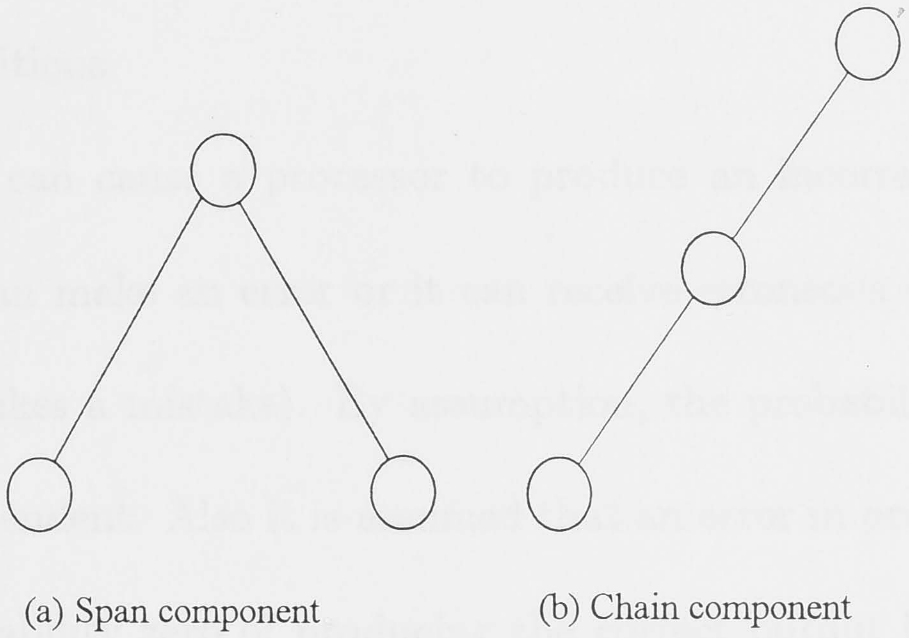


Figure 2.6:

The span and chain components, shown in Figure 2.6, are the two smallest hierarchies that organize the same number of people in difference ways (there is only one hierarchy for one processor and only one for two processors). If (everything else equal) spans perform more reliably than chains of command, then it must be the case that the span component in Figure 2.6(a) has a higher overall reliability than the chain component in Figure 2.6(b). This is referred to as *span dominance*. If the converse is true and chains have higher overall reliability than spans, everything else equal, then *chain dominance* is said to hold.

Intuitive as these dominance conditions are for simple hierarchies, trying to compare larger hierarchies by some sort of decomposition into simpler subhierarchies is an impractical method in general. However, Theorem 9 provides a simple method for using the primary reliability function θ to implement the dominance conditions.

Two factors can cause a processor to produce an incorrect output (to be unreliable), it can make an error or it can receive erroneous inputs (one of its subordinates makes a mistake). By assumption, the probabilities of these two events are independent. Also it is assumed that an error in processing incorrect inputs has probability zero of producing the correct output by accident - two wrongs don't make a right. Thus the probability of a processor being correct can be written as the product of its primary reliability times the overall reliability of its inputs. To express this formally, consider a processor P_0 with overall reliability R_0, and I_0 raw data inputs. In addition let P_0 have J subordinates, $\{P_1, ..., P_J\}$, with respective overall reliabilities $\{R_1, ..., R_J\}$. Then

$$R_0 = \theta_n (I_0 + J) \times \prod_{i=1}^{J} R_i. \tag{2.13}$$

Similarly, the overall reliability of P_0 would just be $R_0 = \theta_n (I_0)$ if P_0 had no subordinates. Using recursion and independence allows the overall reliability for a hierarchy G, (denoted $R(G)$) to be calculated directly from the primary reliabilities. Let processor $i = 1, .., P$, have I_i raw data inputs and s_i subordinates, then

$$R(G) = \prod_{i=1}^{P} \theta_n (I_i + s_i). \tag{2.14}$$

Assume, for convenience, that $N = 7x$, so that each processor is assigned x pieces of data. Also let $m(j)$ denote the number of processors with j subordinates, then the expression can be simplified further

$$R(G) = \prod_{j=1}^{M} \theta_n (x + j)^{m(j)}. \tag{2.15}$$

Where M is the greatest number of subordinates of any processor in G.

Equation 2.15 gives the overall reliability of a hierarchy if θ_n is known. In general the functional form of θ_n is not known. However Theorem 9 gives a way to combine Equation 2.15 and our intuition over span and chain dominance in order to rank the overall reliabilities.

Theorem 9 *The reliability function θ_n is strictly log concave on $\mathbf{Z}^+$ (strictly log convex on $\mathbf{Z}^+$) if and only if chain dominance holds (span dominance holds).*

Proof.

By Equation 2.15, the reliabilities for the span and chain components are $\theta_n (x + 2) \theta_n (x)^2$ and $\theta_n (x + 1)^2 \theta_n (x)$ respectively.

If span dominance holds then

$$\theta_n (x + 2) \theta_n (x) > \theta_n (x + 1)^2. \tag{2.16}$$

Taking logs and dividing by 2 gives

$$\frac{\log \theta_n (x+2) + \log \theta_n (x)}{2} > \log \theta_n (x+1). \tag{2.17}$$

Equation 2.17 holds if and only if $\log \theta_n (x)$ is strictly log convex on $\mathbf{Z}^+$. The proof for chain dominance proceeds in a similar manner.

■

Which of span and chain dominance is more appropriate for determining the reliability of a hierarchy for a given decision problem is exogenous to the analysis here. It would depend on the type of decision the hierarchy is making and is an empirical matter.

One consequence of equation 2.15, is that different hierarchies can have the same reliabilities. This occurs because only information on the number of subordinates and raw data inputs is used, which does not fully describe the shape of a hierarchy. As a result this refinement cannot in general guarantee a unique outcome as the following example shows: span dominance to the set of four efficient hierarchies in Figure 2.5.

Using equation 2.15 the reliabilities for the four hierarchies in Figure 2.5 are

$$(a) \quad \theta_n(x+3)\theta_n(x+1)^3\theta_n(x)^3$$

$$(b) \quad \theta_n(x+2)^2\theta_n(x+1)^2\theta_n(x)^3$$

$$(c) \quad \theta_n(x+3)\theta_n(x+2)\theta_n(x+1)\theta_n(x)^4$$

$$(d) \quad \theta_n(x+3)\theta_n(x+2)\theta_n(x+1)\theta_n(x)^4$$

Notice that (c) and (d) have the same reliabilities under this method and hence this refinement will be unable to differentiate between them.

Comparing (a) and (b), using equation 2.16 gives

$$\theta_n(x+2)\theta_n(x) > \theta_n(x+1)^2$$

therefore

$$\theta_n(x+3)\theta_n(x+1) > \theta_n(x+2)^2$$

hence

$$\theta_n(x+3)\theta_n(x+1)^3\theta_n(x)^3 > \theta_n(x+2)^2\theta_n(x+1)^2\theta_n(x)^3$$

and thus (a) is preferred to (b) under span selection.

Similarly for (c) and (a).

$$\theta_n(x+2)\theta_n(x) > \theta_n(x+1)^2$$

hence

$$\theta_n(x+3)\theta_n(x+2)\theta_n(x+1)\theta_n(x)^4 > \theta_n(x+3)\theta_n(x+1)^3\theta_n(x)^3$$

giving (c) preferred to (a). Thus span selection gives $\{(c),(d)\}$ as the reduced set of equilibrium efficient hierarchies. Intuitively these are the hierarchies which appear to feature shorter chains of command.

One advantage of the formalization presented above is that it provides a concrete method for partially ordering the hierarchies that is equally valid for large hierarchies where intuition is a much poorer guide than in the simple example presented here.

2.5.3 Lateness: Time Reliability

An alternative view of stochastic performance by a processor is to consider the possibility of lateness by a processor instead of the possibility of error. A simple model of lateness in production is given in Sobel(1992), where an individual can forget all that they have done so far and have to start again. This can give rise to very long, and unrealistic completion times for information processing tasks. The alternative I suggest, is that an individual can be one period late in passing their output, with probability $(1-\phi_n)$, where again $\phi_n \to 1$ as $n \to \infty$. Similar to the trembling hand refinement in game theory, $(1-\phi_n)$ is infinitesimal. Thus the probability of a processor being more than one period late is the product of infinitesimal numbers and is hence too small to be of concerned. Being a number of periods late would give no extra predictive power and would only complicate the model. Besides it is intuitively unlikely to give a positive probability to an individual being ten years late with a weekly report. In a firm an individual may well be late, perhaps even a number of periods, but it is implausible that an employee would keep their job if such a situation ran on particularly long.

Let the probabilities that different processors are late be independent. The probability that more than one processor is late is again the product of infinitesimal numbers and hence can be ignored in the analysis. Thus we only consider the possibility of lateness occurring at one processor in the hierarchy.

Lateness by a processor has two possible affects. It can increase the delay in making the final decision by the hierarchy and it can cause processors to be active for extra periods of time. Clearly if a processor is late then it must have worked an extra period, it will also cause processors at higher levels to work extra periods if they were relying on its output being made on time (either directly or indirectly). We focus here on the extra periods of work because it gives a higher level of differentiation between hierarchies (that is it is more likely to give a unique outcome).

The maximum extra delay due to lateness is either 0 or 1. It will be 0 if at some point above the late processor, some other processor has inputs such that it can organize its activities to overcome the lateness. The extra delay will be 1 otherwise. An example of such a rescheduling would occur if one of the single processors (with no subordinates) immediately below the top processor in figure 2.5(c) was late. In this case there are two processors passing reports at the same time, thus if one is late the top processor just deals with the other first. Thus focusing on extra delay does not take account of the shape of a hierarchy but just the occurrence of this rescheduling possibility at some point.

Focusing on extra periods worked captures both structural elements of the hierarchy and the rescheduling potential. This occurs because lateness by a processor will cause all processors above it to work an extra period until a processor with the potential to reschedule is reached. The hierarchy(s) with the lowest expected extra periods of work is preferred. Thus this refinement will favor shorter chains of command and the rescheduling potential.

The potential to reschedule only exists if there are some reports which are not immediately processed by the person receiving them. This is exactly the opposite of the just-in-time message condition for MS networks in Van Zandt(1997c). Van Zandt uses MS networks as a characterization of a class of one shot efficient networks and, more importantly, as the building blocks for a class of efficient periodic networks. However as just observed, MS networks may not be preferred in a stochastic performance setting. This indicates one way in which the results from the existing deterministic theory of information processing networks may be substantially effected by considering a stochastic performance case.

The expected extra periods of work (if only one processor is late) for the four hierarchies in Figure 2.5 are (a) $(1-\phi_n)17$, (b) $(1-\phi_n)18$, (c) $(1-\phi_n)15$ and (d)$(1-\phi_n)12$. Thus under the lateness refinement (d) is the unique equilibrium hierarchy.

2.6 Conclusion

The major innovation of this chapter has been to use hours as well as number of employees in analyzing information processing hierarchies. Although this extension seems obvious it has not previously been made. Theorem 4 shows, surprisingly, that there exist equivalences between the efficient sets of hierarchies under the processor and piece-rate regimes. However as the remainder of the chapter shows, analysis at the hours level can give new insightful results. In particular, constant returns to scale were shown if hours were used instead of employee numbers.

The methodology of looking at individual activity is also the driving force behind the refinements developed in Section 2.5. The refinements significantly increase the number of situations in which unique outcomes are predicted. Uniqueness then opens up the potential for empirical analysis of firm organizational charts and the decision making environment they operate in.

Finally one of the most attractive areas of future research is in the synthesis of the information processing approach with the principal agent/incentive literature. If incentive considerations are to be introduced in determining the employment contract then performance (specifically θ and ϕ) should depend on effort in someway, and effort should be related to the number of periods worked. Again this will rely on the analysis of individual activity and stochastic performance introduced in this chapter.

Appendix

Proof. Lemma 7.

Levels in a hierarchy are defined as follows. Processors with no subordinates are of level 1. The level of any other processor is 1+ the maximum of the levels of its subordinates.

By efficiency, every level one processor receives raw information, otherwise

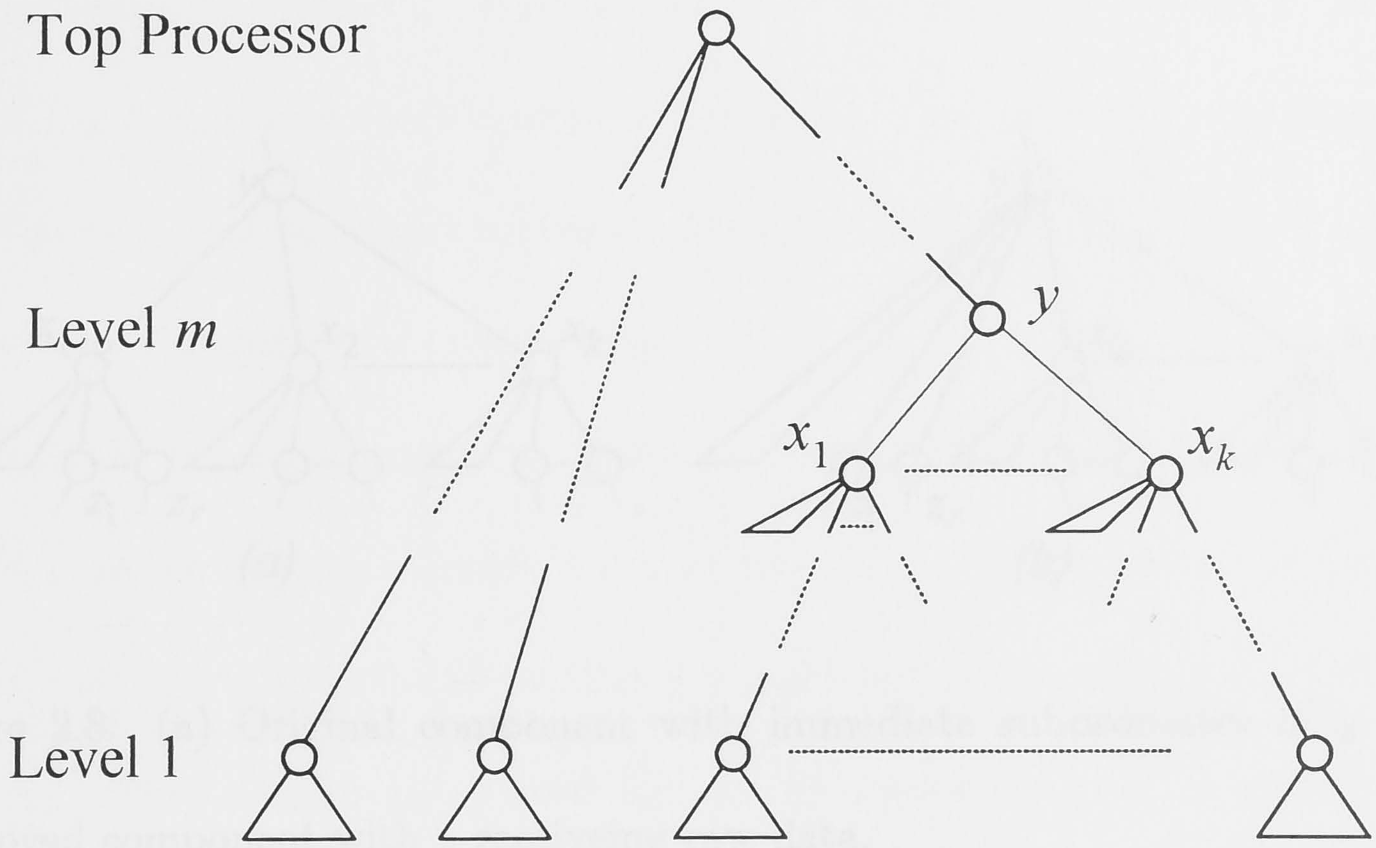


Figure 2.7: Hierarchy with raw information only received below level m.

it would do nothing. The result follows immediately for single level hierarchies, thus we now consider multiple level hierarchies.

For the purpose of strong induction assume that all processors below level $m > 1$ receive raw data. For an example of such a hierarchy see Figure 2.7.

Consider a level m processor y which receives input only from the processors $\{x_1, ..., x_k\}$. Such a processor, y, from Figure 2.7 is shown in Figure 2.8(a), with its immediate subordinates. A part or subsection of a hierarchy, such as that shown in Figure 2.8(a) is referred to as a *component*. The timing diagrams for these two components are shown in Figure 2.9 (a) and (b), respectively.

Assume without loss of generality that the subordinates of y pass their reports in numerical order according to their index (x_1 first, and so on). Let x_1

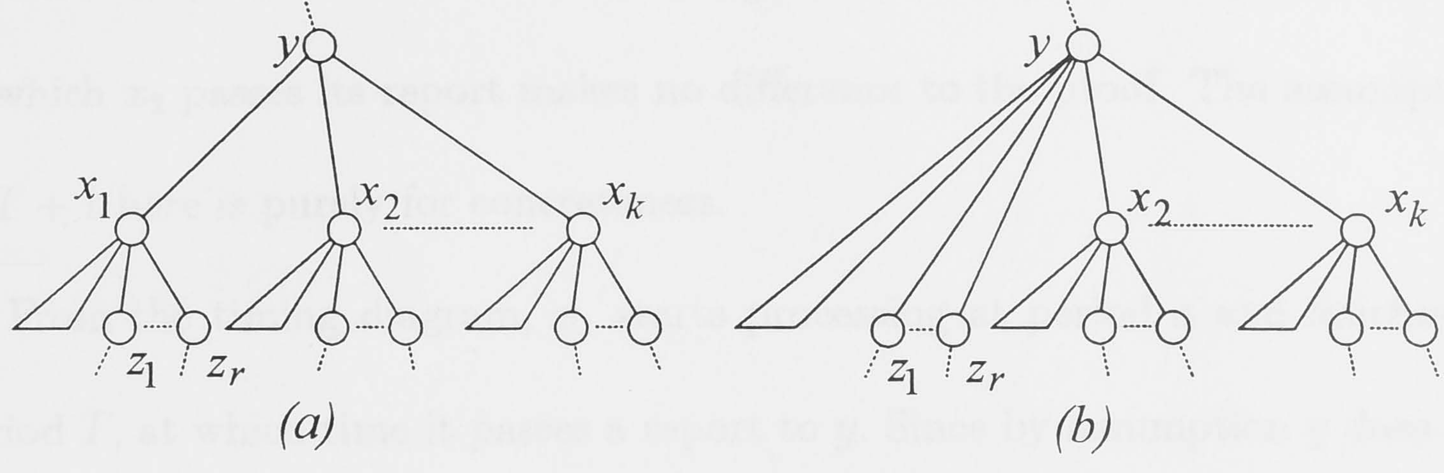


Figure 2.8: (a) Original component with immediate subordinates of y. (b) Improved component with y receiveing raw data.

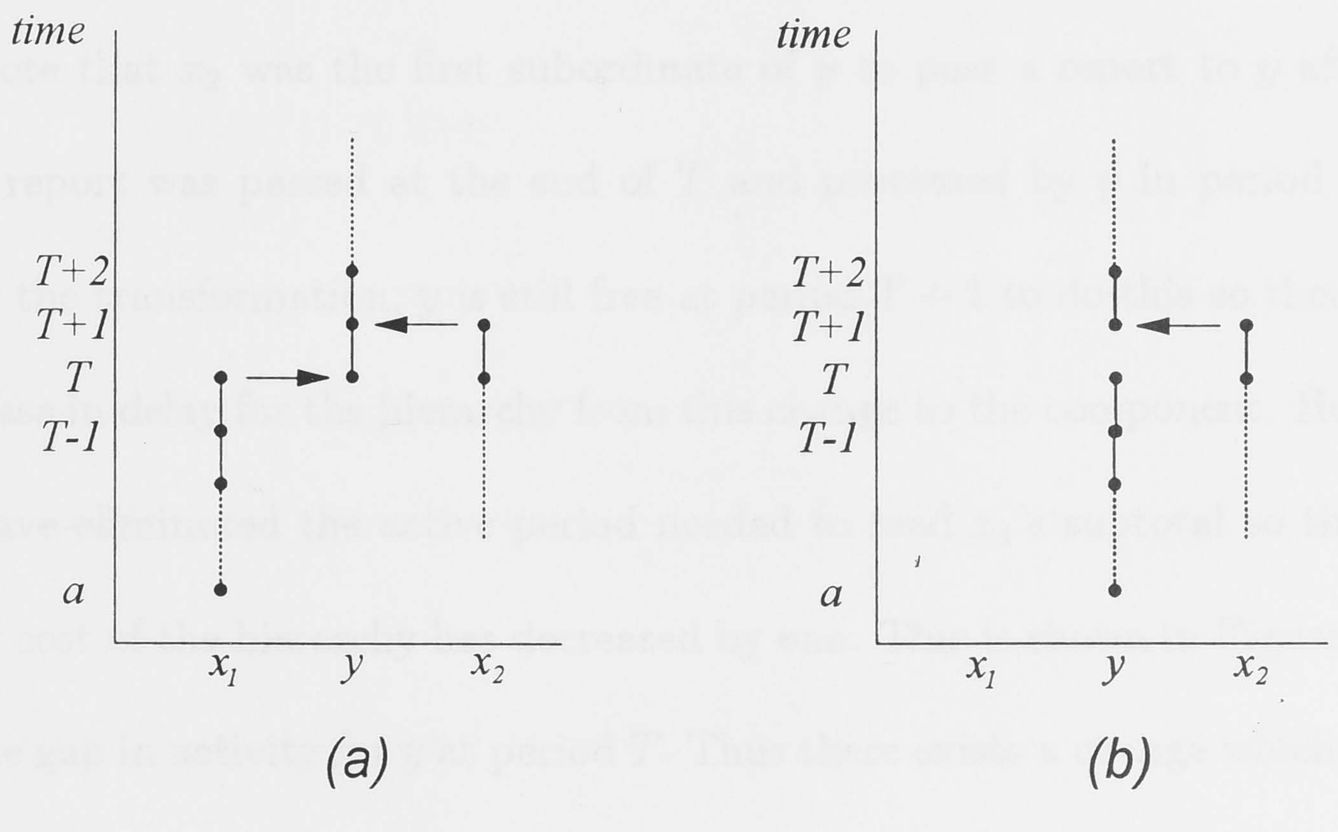


Figure 2.9: Timing diagrams: (a) for original component ; (b) for improved component.

pass its report in period T, then the earliest x_2 could pass its report to y is period $T+1$, this is the case shown in Figure 2.9. The particular period after T in which x_2 passes its report makes no difference to the proof. The assumption of $T+1$ here is purely for concreteness.

From the timing diagram, x_1 starts processing at period a and finishes at period T, at which time it passes a report to y. Since by assumption y does not receive raw information it follows that y is inactive before receiving the subtotal from x_1. Thus x_1 can be deleted and all its inputs directed to y, this new component is shown in Figure 2.8(b). The information can be processed by y in the same time as $x_1$used to take, hence y can also start at period a and finish the work at period T that x_1 used to perform.

Note that x_2 was the first subordinate of y to pass a report to y after x_1. This report was passed at the end of T and processed by y in period $T+1$. After the transformation, y is still free at period $T+1$ to do this so there is no increase in delay for the hierarchy from this change to the component. However, we have eliminated the active period needed to read x_1's subtotal so the total labor cost of the hierarchy has decreased by one. This is shown in Figure 2.9(b) by the gap in activity for y at period T. Thus there exists a change which can be made to the original hierarchy which reduces the quantity of labor used without increasing the delay. Hence the old arrangement cannot have been efficient. By assumption x_1 received raw data so y now receives raw data as well. Hence by induction all processors at level $m > 0$ receive raw data. ■

Chapter 3

A Monopoly Location Problem with Hierarchies

3.1 Introduction

This chapter develops a model of management structure and function for a firm which is developing a product for a predetermined launch date. The example which typifies the situation being considered is the development of a feature film.

There are standard distribution deals for many films and in the short term the price that consumers pay for different films remains fixed. Hence film producers compete not on price but the characteristics of their films. The preferences of the film going public are not directly observable and change over time. However market research can and does influence the content of films during the production process.

This chapter attempts to address two main questions motivated by this example. What kind of management structure should a firm in this kind of market adopt in order to try and maximize profits? What effects do differing management structures have on performance?

The product is one that can vary in its characteristics, hence the firm has to pick a combination of characteristics for its product which it believes will maximize returns. The preferences of consumers change over time and are not directly observable. To describe this situation a standard model of one dimensional preferences is expanded to a dynamic stochastic setting where a firm consists of boundedly rational agents. In particular, agents are limited in the amount of computation they can perform in each time period.

As a result of the stochastic nature of the problem, a firm is forced to make a decision on what product to produce, based on a forecast of the unknown market conditions at the launch date. Since the firm has no competitors it simply decides to locate at which ever point its forecast tells it will be the centre of the market when the product is launched.

It is shown in Radner(1993) and Van Zandt(1994), that the time to calculate a forecast is an increasing function of the amount of data considered and a decreasing, bounded function of the number of people employed in making the decision. By applying earlier work on hierarchies to the stated economic problem, this chapter has succeeded in endogenizing the cost of delay (the time taken to calculate the forecast).

Each decision making procedure (method for producing a forecast) corresponds to a different management structure and associated management function, which are modeled here as information processing hierarchies. This methodology allows analysis of the minimum delay incurred by a given organizational structure in making a decision based on a given quantity of market research.

The principal intuition of the chapter is that increasing sample size gives a more accurate estimate of a historical situation, but it also increases delay, making this estimate a more out of date basis for predicting the future. In formalizing this intuition it is also shown that the optimal size of a firm's management structure depends on the speed of change that occurs in the product market. The relationship between the type of product sold, organizational structure and expected profit is also examined.

Section 3.2 formally defines the model which will be considered. Section 3.3 examines the properties of the single sample case in some detail. A mixture of numerical and analytic techniques are used. Section 3.5 is the conclusion.

3.2 The Formal Model

3.2.1 The Market

The traditional location model of differentiated products is modified to become a dynamic stochastic model. We begin by describing the situation during period

t, and then present the process by which market conditions change over time.

Consumers derive satisfaction from the consumption of the intrinsic characteristics of the goods they purchase. For convenience we assume that the good under consideration has one relevant characteristic, which can be represented by a real number. A good produced by a firm can be represented by the unique point x on the real line (the real line is referred to as the characteristic space). For ease of exposition, a good is generally identified with the point representing its characteristics, so that the good itself is said to be "located" at the point on the real line.

Individuals differ in their preference for characteristics. Each individual has a unique ideal type of good, which is also represented by a point, c, also on the real line. Consumers prefer products that are closer to their ideal points. In particular, an individual's utility from a product is strictly decreasing in the distance between the product and the individual in the characteristic space (similar to goods we refer to consumers as being located in the characteristic space at their ideal point).

Individuals differ in their location and are represented as a continuum of consumers, whose locations are uniformly distributed over the interval $[M(t) - \frac{1}{2}, M(t) + \frac{1}{2}]$, referred to as the *market interval.* Consumers have unit demand and buy whichever product gives them the highest utility, which will be the closest product to their ideal point. They only buy if the utility from a product satisfies some reservation utility constraint (which would be determined exogenously by outside options).

We assume that the reservation utility level u^* is the same for each consumer and that the utility functions are identical functions of the distance of a product from an individual's ideal point. The utility for a consumer with ideal point c, from good x is given by

$$U(x,c) = V(|x-c|) \tag{3.1}$$

where $V(|x-c|)$ is strictly decreasing in $|x-c|$. Let $z = x - c$.

Notice that the utility function of each individual is symmetric about that individual's ideal point. Much of the following could be generalized to non-symmetric utility functions. The functional form in Equation 3.1 is used for its expositional ease, and because the main focus of this chapter is to elucidate the inter-relationship between internal information processing and market behavior in as general and concise a way as possible. The complexities of more specific functional forms may well be needed when this model is fitted to data.

Fixing the price reduces the generality of the model in one sense but increases it in another. In order to allow price to vary in a location model it is necessary to explicitly define the relationship between price and spatial component of utility. Results derived from such a specification are often dependent on the exact functional form chosen.[1] By fixing price here, choice of a particular functional

[1] Such an example is discussed in Tirole (1988, p280-281). In the fixed location, linear cost model no pure strategy Nash equilibrium in prices exists when the firms are located sufficiently close to the centre of the market. However in the same model with quadratic transport costs

form (such as linear transport cost) is avoided, leading to results which are more general.

In the rest of the chapter we will consider a firm in isolation. The firm will have buyers whose ideal points satisfy $V(z) \geq u^{*}$. Since the utility function is strictly decreasing and symmetric, it follows that there exists some $r \geq 0$, such that consumer c will only purchase good x if $|x-c| \leq r$. Hence the potential customers for product x are those that lie in the closed interval $[x-r, x+r]$, referred to as the *product interval of appeal*, or the *firm's interval* (as opposed to the market interval).

Consumers are represented by a probability distribution (with density $h(c)$) over the characteristic space. Integrating the density function over product x's interval of appeal will give the proportion of the total population of consumers who would buy product x. Since consumers are uniformly distributed on $[M(t)-\frac{1}{2}, M(t)+\frac{1}{2}]$, the proportion of consumers purchasing x (referred to as *market share* $s(x, M(t), w)$) is given by

$$s(x, M(t), w) = \int_{x-r}^{x+r} h(c)dc = m([x-\frac{w}{2}, x+\frac{w}{2}] \cap [M(t)-\frac{1}{2}, M(t)+\frac{1}{2}]). \quad (3.2)$$

Where $m(.)$ is the measure of a set (which is simply the length on the overlap in this case) and $w = 2r$.

It should be noted that this is a share of the potential market (which is the whole population), not a firm's share of actual sales in the market. Hence this

---

a pure strategy Nash equilibrium in prices does exist.

definition of market share differs from the one in common use. By definition a monopolist's share of sales is always 100%, which is completely uninformative.

The midpoint $M(t)$ of the consumer distribution is assumed to be a function of time. For expositional ease we assume that $M(t)$ a random walk random walk:

$$M(t) = M(t-1) + \varepsilon_t, \tag{3.3}$$

with $\varepsilon(t) \sim N(0, \sigma^2)$.

3.2.2 The Firm

A firm has to decide where to locate its product in order to maximize expected profits. The launch date is assumed to be fixed in advance. The firm cannot observe the market directly but it can attempt to locate the market by surveying consumers. The management problem is to attempt to maximize profits by choosing when and how much to sample, a time frame for processing the samples and an organizational structure with which to process the samples. These choices are not independent and are determined by the production and management technology available to the firm.

The technological situation a firm faces is considered next.

Decision making by the Firm

This is a model of product development for a fixed launch date in the face of evolving consumer preferences. Firms only gain information about what is happening in the market by surveying consumers. The product location decision (i.e. which x to produce) must be made prior to launch so there is no sales data from the product on which to base inference about the market. Instead a firm must go out and conduct some form of market research if it is to locate and track the population of consumers in the characteristic space. This market research might be in the form of questionnaire surveys or product pre-release tests. For example, feature films often have test screenings and their endings or offensive scenes are changed depending on viewer reactions. Another example is that some producers of computer software distribute beta versions of programs to test out new features before release.

In order to capture as large a market share as possible, a firm really needs to know about the shape and location of the market when it releases its product. Conceivably, firms might be using market research to infer the distribution of consumer preferences, the form of the process by which this distribution changes over time and the history of shocks that have a persistent influence on the location of the market. Similar to most of the literature on a monopolist learning about its demand conditions, we simplify the problem the firm faces by assuming that it knows all the relevant functional forms and the values of all coefficients and variables, except the history of actual outcomes of $M(t)$'s and the $\varepsilon(t)$'s, which are the shocks which occur to the market.[2]

Firms will be able to guess the correct location of $M(t+1)$ in period t with

[2]For an example of learning in a differentiated products market see Harrington(1995)

probability 0 because there is a white noise shock term. However, a firm will get a positive payoff as long as its product interval overlaps with the market. The market transition process, Equation 3.3, is autoregressive, hence knowledge about past locations of the market will be useful in producing an estimate of where the market will be located at launch time.

A firm is interested in acting to maximize expected profit. The profit from launching product x in period t is denoted by π. The profit will be the revenue, R, generated from the market share the product captures, less the costs. Costs are of two types: a constant marginal cost of production c and a fixed cost, F.

Assume that price p and the size of the market Q are constants. Then the quantity sold is $Qs(x, M(t), w)$, and $R = pQs(x, M(t), w)$. Hence expected profit is given by

$$\begin{aligned} E[\pi] &= E[R - cQs - F] \qquad (3.4) \\ &= (p-c)QE[s(x, M(t), w)] - F. \end{aligned}$$

Potentially there are two sources of fixed cost: fixed costs in the production of the good and management overheads. In this model, management overheads could be of two types: the cost of collecting information and the labor cost of processing that information in order to make a decision. We are concerned here with how internal structure and the behavior of firms are related, hence the labor cost of processing the information is of more interest. However, no method for endogenizing wages (and hence labor cost) has yet been devised for this type of model. An arbitrary choice of wages offers little insight, so the whole question of labor cost has been left as an avenue of future research. Thus the approach used here is similar to Van Zandt and Radner(1995) Van Zandt(1997d).

The limited capabilities of a firm's employees are explicitly modelled by describing each individual as a processors and the organization as a programmed network.

A processor has an *inbox* and a *register* (which together comprise its *memory*). In one period of time a processor can perform one of:

1. Read one number from the sample or its inbox and store it in its register, overwriting the previous contents of the register.

2. Read one number, y, from its inbox, calculate a linear function of y and the contents of its register and store the result in its register;

In addition to one of the above operations a processor can, at the end of a period, instantaneously write the contents of its register to the inbox of any processor to which it is connected (this is called *output*).

Together processors and connections form a *network*. A *program* specifies which operations each processor performs in each period.

A *hierarchy* is a network that contains no loops or cycles of connections. Thus in a hierarchy information never comes back to a processor once it has been outputted. Supervisors and subordinates are defined for a hierarchy as follows. A *supervisor* of a processor is the processor to whom output is passed.

The *subordinate(s)* of a processor are the processor(s) from whom output is received. Processors with no subordinates receive only data from the outside world and are defined to be at level one. If a processor has subordinates then its level is one more than the maximum level of its subordinates. There is a unique processor, called the *top processor* or the *boss*, which has the highest level in the hierarchy.

There is only one decision to be made by the network: which product to make. It is shown in Radner(1993) that for this type of decision and a network of processors, it is sufficient to consider only hierarchies. To describe the time taken to make the decision we introduce the concept of delay. Delay is the amount of time a hierarchy takes from the arrival of the first piece information to be used until the decision is produced.

Firms certainly use information about market conditions when they launch a product. This information can rarely, if ever, be on every consumer, hence either explicitly or implicitly firms must use statistical inference based on what is essentially sample information to predict for the whole population of consumers. As a result the language of statistical sampling and market research will be used to describe the way in which the firm uses information.

The model should not be thought of as a literal attempt to describe the process of market research, rather the model tries to capture the interactions of information, organizational structure and product market decisions.

Although specified here as a linear function, so that the network can calculate a sample mean, the generalized form of this model only assumes that processors can perform an associative binary operation. The adding of numbers should be considered as an analogy of the processing of information and the producing of reports, recommendations and decisions that really go on inside the management structure of an organization. It might seem a trivial task to collate sample information, but the information which flows within this programmed network should also be thought of as the decisions and recommendations that are based on raw information.

Section 3.3 examines the case of single period sampling in some detail. Numerical techniques are used to determine the relationship between the information processing structure, expected market share and the type of product.

3.3 The Single Sample Approach

Potentially a firm might choose any number of samples and any decision rule based on these in order to determine where to locate its product.

The simplest possible approach is to take no sample at all and locate at the point which the entrepreneur believes will maximize profits. If the process by which consumer preferences change is stationary, then there exists a long run mean to preferences and this would be the sensible place at which to locate if no other information is used. This approach is likely to do well if the shifts in the market are small compared to the width of appeal of the product. This approach makes a useful bench mark for comparison in stationary markets.

If, as assumed here, the process by which the mid point of the market shifts is non-stationary, then by definition there does not exist a mean for the market transformation process which is independent of time. Thus there is no sensible reason for picking any point without sample information, since the expected profit for a randomly drawn location is zero.

The next simplest case is to take a single sample and to make all decisions based upon this. Although perhaps not the optimal sampling approach if sampling is costless and unconstrained, this approach is intuitively plausible.

Sampling is not in general free and may have sizable fixed or marginal costs limiting the size and number of surveys taken. In addition, there are institutional factors that may well constrain the number of surveys. The surveying process has not been modelled here, it just provides a number of data points which are to be processed to make a decision. The most common paradigm for market research is some kind of questionnaire based survey. It is however important to consider whether the survey questions are general or based on responses to a specific experience.

Returning to the feature film example, a general question such as "How much violence do you like in a film?" is going to be of limited use to film producers trying to decide if their film has too much violence. Instead they would like to know, in response to seeing their film, how many people thought it contained too much violence, and which particular scenes were disturbing. In order to gain this information pre-release showings have to be arranged. It is not just the direct cost of re-editing and printing the film and organizing a showing that make these showings rare. It is important for products, like films, to hit the market with an up beat bang. Too much pretesting can indicate to consumers a product with a problem, the product looses its novelty and it gives more opportunity for competitors to copy or duplicate the product.

For these reasons a firm may use only one sample, and then just decide how best to process that sample, and simultaneously with the choice of processing, how large the sample should be. This situation is close to the sampling problem of classical statistics, but in this model the cost associated with a sample is now endogenously determined by its impact on profits, rather than the cost of actually collecting the data.

We begin by assuming that the firm takes a sample of size N which is the only source of empirical information used in the product decision rule. This sample is used to calculate a sample mean and the firm relocates on the basis of this sample mean. The calculation of the sample mean is performed by the people who work for the organization, by forming some programed network.

boss at level L) and that individuals at all levels above 1 have subordinates, the number of subordinates is known as an individual's span of control. The second assumption is that either all individuals at the same level have the same span of control (hence reducing the hierarchy to L parameters) or that the span of control at each level is the same, reducing the hierarchy to two parameters.

In a binary hierarchy all spans of control are two. The binary hierarchy example is useful because its hierarchies are easy to visualize[3]. The functions relating delay and sample size are also well behaved and can be parameterized by the single variable L.

In a binary hierarchy there is one person at level L, 2 people at level $L-1$, and so on down so that there are 2^{L-1} employees at the bottom (level 1). Summing over the number of levels gives $\sum_{i=1}^{L} 2^{i-1}$ total individuals in the hierarchy.

It takes an individual two periods to add up two pieces of information, regardless of whether they are subtotals received from another person in the hierarchy or raw data from the sample. Assuming that each individual at level 1 receives 2 pieces of raw sample data, then the hierarchy receives $N(L) = 2^L$sample points and takes $D(L) = 2L$ periods to process them into a decision.

Figure ?? gives an example of a three level binary hierarchy processing the sample mean of the data set: $\{C1, ..., C8\}$. Individuals are represented by the boxes. Each box shows the linear computation performed by that individual. The data points are shown being read in at the bottom of the hierarchy.

[3]Unconstrained choice of hierarchy is considered in the next chapter.

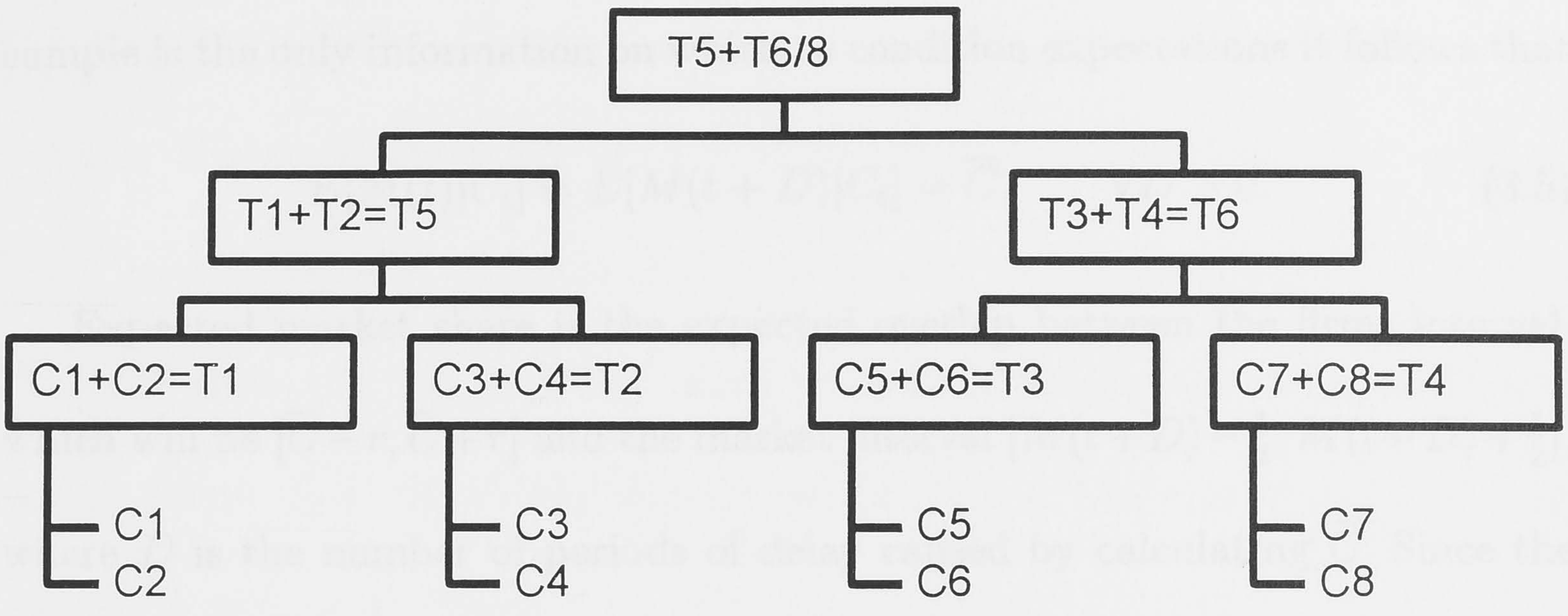


Figure 3.1: A three level binary hierarchy calculating the sample mean of the set {C1, ...,C8} of observations. Ti denotes a subtotal.

These assumptions reduce a firm's organizational and sample decision to a single parameter, L. By choosing the number of levels in the hierarchy a firm also fixes the amount of information it can process and the amount of time it takes to do this processing and hence in which period to take the sample (since for equal sized samples, the most recent one will always be the most useful).

Next we need to consider the expected market share and hence expected profits associated with each size of hierarchy, and how these depend on the environment in which the firm operates.

Assume that the survey was taken in period t. The sample $C_t = \{c_1, \ldots c_N\}$ that a firm receives is the set of ideal points for the consumers who were surveyed. By calculating the sample mean $\overline{C}$, the firm has an unbiased estimate of $M(t)$, which was the mean of the distribution from which the sample came. Since this sample is the only information on which to condition expectations it follows that

$$E[M(t)|C_t] = E[M(t+D)|C_t] = \overline{C}, \qquad \forall D > 0. \tag{3.5}$$

Expected market share is the expected overlap between the firms interval, which will be $[\overline{C}-r, \overline{C}+r]$ and the market interval $[M(t+D)-\frac{1}{2}, M(t+D)+\frac{1}{2}]$, where D is the number of periods of delay caused by calculating $\overline{C}$. Since the market shifts according to a random walk with Normally distributed shocks, the market interval can be rewritten as

$$\begin{aligned} & [M(t+D)-\frac{1}{2}, M(t+D)+\frac{1}{2}] && (3.6)\\ = \; & [M(t)+\sum_{i=1}^{D}\varepsilon(t+i)-\frac{1}{2}, M(t)+\sum_{i=1}^{D}\varepsilon(t+i)+\frac{1}{2}] && (3.7)\\ = \; & [M(t)+T-\frac{1}{2}, M(t)+T+\frac{1}{2}]. && (3.8)\end{aligned}$$

Where $T = \sum_{i=1}^{D}\varepsilon(t+i)$, is the sum of the shock terms. The shock terms are independent and identically distributed, $\varepsilon(t) \sim N(0,\sigma^2)$, hence $T \sim N(0, D\sigma^2)$.

The length of the overlap is a function of $z = M(t)+T-\overline{C}$, the distance between the centres of the two intervals. There are two cases depending on the size of w the firm's interval width. The following analysis assumes $w \leq 1$, this is the economically more interesting case of a product whose width of appeal is no bigger than the market.

$$s(\overline{C}, M(t)+T, w) = \begin{cases} w & if \quad 0 < |z| < \frac{1-w}{2} \\ \frac{1+w}{2} - |z| & if \quad \frac{1-w}{2} < |z| < \frac{1+w}{2} \\ 0 & if \quad \frac{1+w}{2} < |z| \end{cases} \tag{3.9}$$

To find the expected overlap the distribution of $z = M(t) + T - \overline{C}$ is needed. Let $g(z)$ be the probability densities of z.

Now $E[z] = E[M(t) + T - \overline{C}] = 0$ and $Var[z] = Var[M(t) + T - \overline{C}] = Var[T] + Var[\overline{C}]$. The variance of T is the sum of variances of the shock terms, of which there are $2L$, and the variance of $\overline{C}$ is simply the variance of the sample mean of 2^L randomly chosen points from a Uniform distribution on $[M(t) - \frac{1}{2}, M(t) + \frac{1}{2}]$. Hence

$$Var[T] + Var[\overline{C}] = 2L\sigma^2 + \frac{1}{12 \times 2^L}. \tag{3.10}$$

Although $g(z)$ is not known, it can be approximated by its asymptotic distribution, which is Normal. This gives the approximation: $z \sim N(0, 2L\sigma^2 + \frac{1}{12 \times 2^L})$, where the density function $f(z)$ of the Normal distribution with mean zero and variance $2L\sigma^2 + \frac{1}{12 \times 2^L}$ is given by

$$f(z) = \frac{1}{\sqrt{2\pi(D(L)\sigma^2 + \frac{1}{12N(L)})}} \exp\left(\frac{-z^2}{2(D(L)\sigma^2 + \frac{1}{12N(L)})}\right) \tag{3.11}$$

$$= \frac{1}{\sqrt{2\pi(2L\sigma^2 + \frac{1}{12 \times 2^L})}} \exp\left(\frac{-z^2}{2(2L\sigma^2 + \frac{1}{12 \times 2^L})}\right) \tag{3.12}$$

Thus, expected market share for $w \leq 1$ is

$$E[s(x, M(t), w)|\sigma, L] = \int_0^{\frac{1-w}{2}} 2wf(z)dz + \int_{\frac{1-w}{2}}^{\frac{1+w}{2}} 2(\frac{1+w}{2} - z)f(z)dz \tag{3.13}$$

This integral equation cannot be solved analytically. However the relationships between the variables can be examined graphically by solving the equation numerically, which is the approach adopted below.

Approximating $g(z)$ with $f(z)$ matches the first two moments for a small sample. It is shown, in the appendix to this chapter, that the true distribution $g(z)$ is symmetric. The Normal distribution is also symmetric so, in fact, the first three moments are matched. However it is also shown in the appendix that the fourth moments of the two distributions differ, so that the kurtosis (fatness of the tails) is inaccurate in the Normal approximation.

An alternative to approximating the true distribution analytically would be to simulate the true distribution using a Monte Carlo method. This is the approach used in the next chapter. The two methods give slightly different numerical values for expected revenue, but qualitatively all of the following results are supported by either method (a comparison showing this is made in section 4.9). Although less accurate, the normal approximation has the advantage of giving a functional form which is useful for developing intuitive explanations in the next section.

3.4 Results

A difficulty arises in that market share is a function of three parameters, L, w and σ. Thus at least one of these parameters must be assigned a specific value in order to produce a graph. We begin by comparing two hierarchies with fixed numbers of levels, L. The cases $L = 1$ and $L = 6$ are shown in Figure 3.2 and Figure 3.3 respectively. These cases are of interest because they correspond to a single individual processing the sample ($L = 1$) and a rather deep hierarchy ($L = 6$). Many large companies have around six layers to their hierarchies, so by real world standards this type of hierarchy is on the big side in terms of levels (although the number of employees here is relatively small because only one task is being considered). The graphs show the market share these hierarchies expect to gain for ranges of w (the width of appeal of the firms product) and σ (the standard deviation of $\varepsilon(t)$).

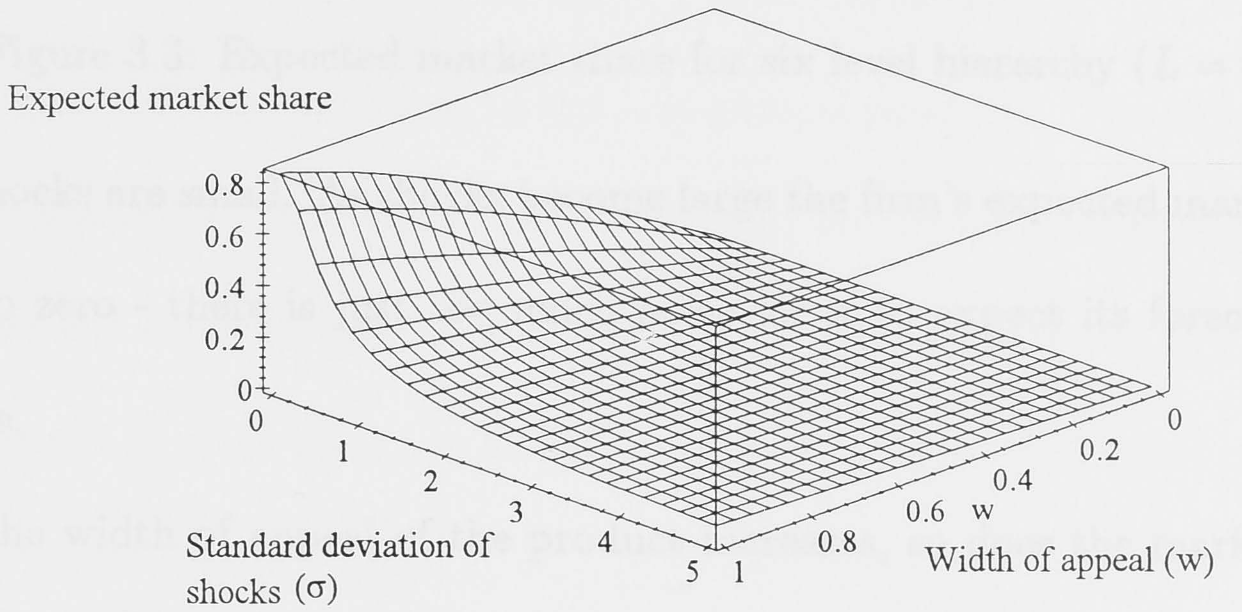


Figure 3.2: Expected market share for one level hierarchy $L = 1$.

Both figures show the same general shape. In both cases a larger market share is obtained when the standard deviation of the shocks to preferences are small. This occurs because when shocks are small there will be little difference between preferences in the period sampled and the period of the launch. If shocks are small, it is not so much speed as accuracy in forecasting which is important, hence the six level hierarchy which takes a larger sample does better

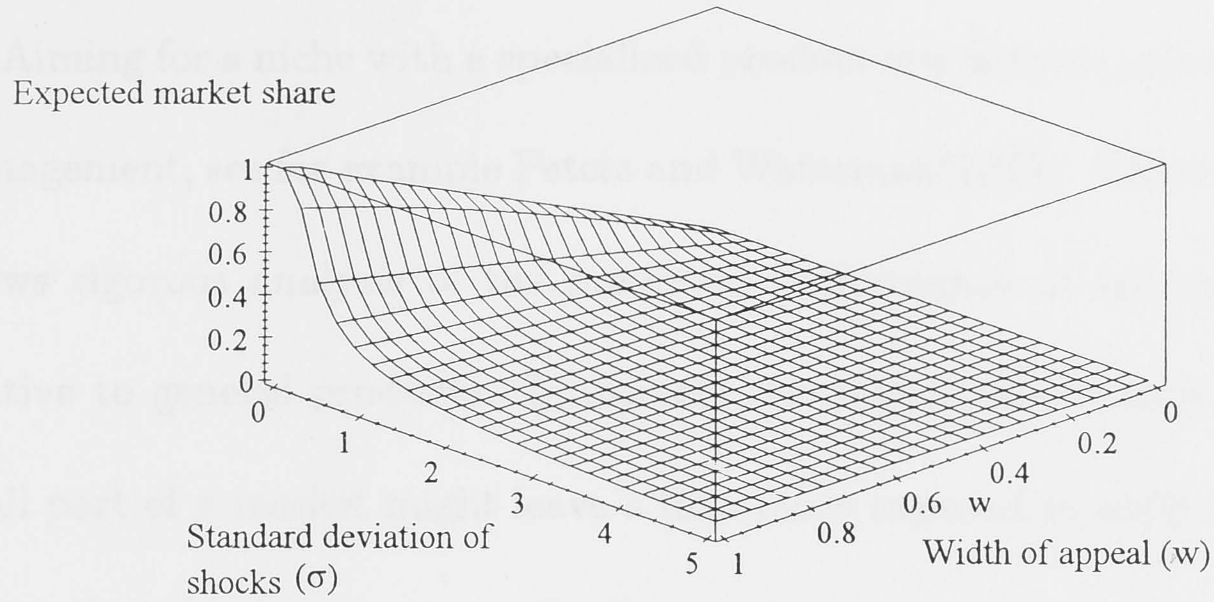


Figure 3.3: Expected market share for six level hierarchy ($L = 6$).

when shocks are small. As shocks become large the firm's expected market share drops to zero - there is just too much noise for it to expect its forecast to be accurate.

As the width of appeal of the product increases, so does the market share, as long as the shocks are not too large. Again, this is intuitive since the market share a firm gains can be no larger than the width of appeal of its product.

The shapes of the two graphs do differ, suggesting that there are potentially interesting interaction effects amongst the parameters. These are not clear from the three dimensional plots. Hence, below, two dimensional graphs for two different values of a third parameter are compared on the same plot to make the effects clearer. We shall focus in particular on two types of products which we shall now define. A *specialized product* ($w = 0.1$), which can gain at most 10% of the market and a *general product* ($w = 1$) which could gain 100% of the market.

Aiming for a niche with a specialized product is a favorite principle of modern management, see for example Peters and Waterman(1982). The model used here allows rigorous analysis of the predicted performance of specialized products relative to general products. Intuitively one might believe that only having a small part of a market might leave a firm more exposed to shifts in preferences; we see that this is not necessarily the case.

First we compare the performance of a small hierarchy ($L = 1$) and a large hierarchy ($L = 6$), for both specialized and general products. The differences in expected market share for these two hierarchies are shown in Figure 3.4. Each line shows the expected market share of a one level hierarchy less the expected market share of a six level hierarchy, for a specific type of product and for a range of shocks.

In both cases the large hierarchy does best when the shocks are small (negative values on the vertical axis) and the smaller (and hence faster hierarchy) does better for larger shocks. Note that, counter intuitively, the specialized product is less affected in both absolute and relative terms by the size of the shocks. The specialized product width is smaller than the market interval, hence by aiming for the middle of the market, the specialized firm gives itself plenty of room to miss. The general product has a wider interval so it always has a larger market share than the specialized firm, but is consequently relatively more affected when its forecast is inaccurate.

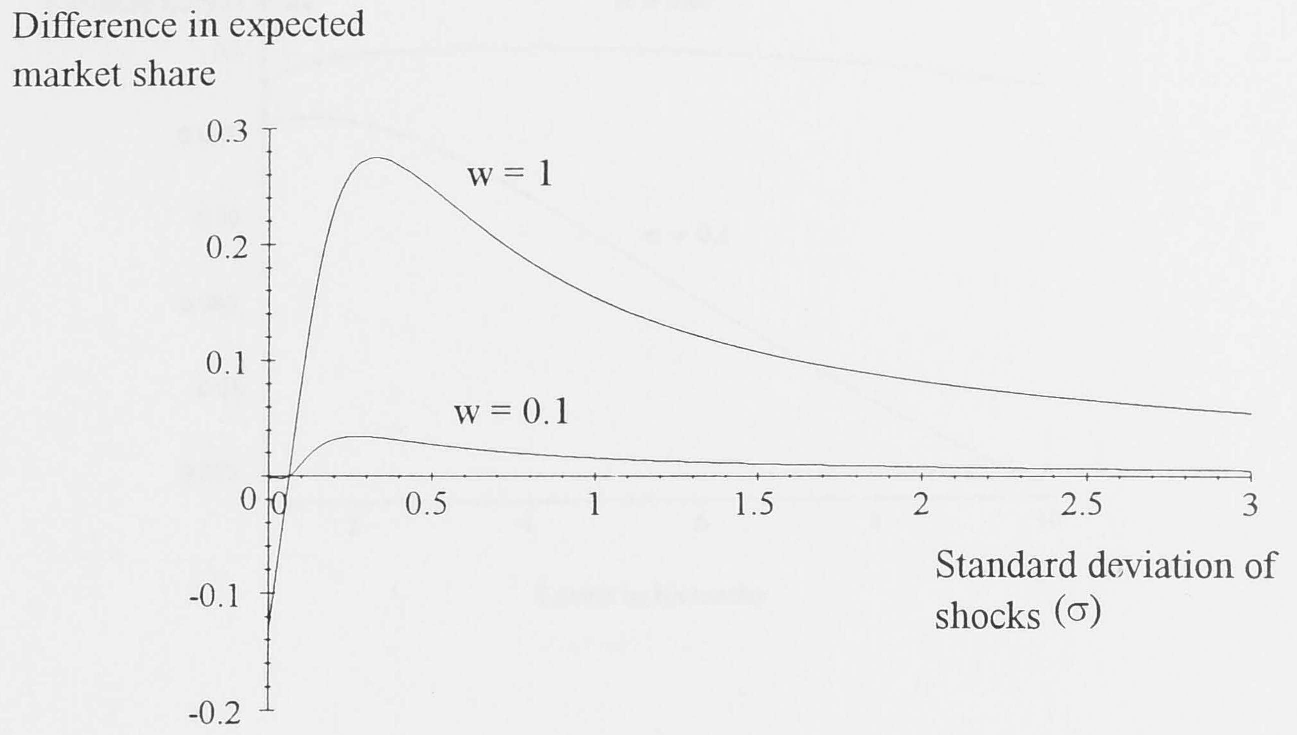


Figure 3.4: Expected market share for a one level hierarchy less expected market share for a six level hierarchy.

Next, in Figure 3.5 we examine the optimal choice of hierarchy for a specialized product when shocks are small ($\sigma = 0.05$) or a little larger ($\sigma = 0.1$). For small shocks the optimal hierarchy has four levels. As the shocks become larger, the optimal size of hierarchy decreases (for the $\sigma = 0.1$ case, it has dropped to a two level hierarchy). When shocks are smaller, the cost of having too large a hierarchy (or too small) are relatively low, shown by the flatness of the $\sigma = 0.05$ line. As the world becomes more uncertain it becomes more important to be the right size. This sits well with the intuition that firms in a very static environment can survive despite inefficient bureaucracies.

The situation for a general product, shown in Figure 3.6, is much the same. The general product is less effected by the increase in the size of the shocks,

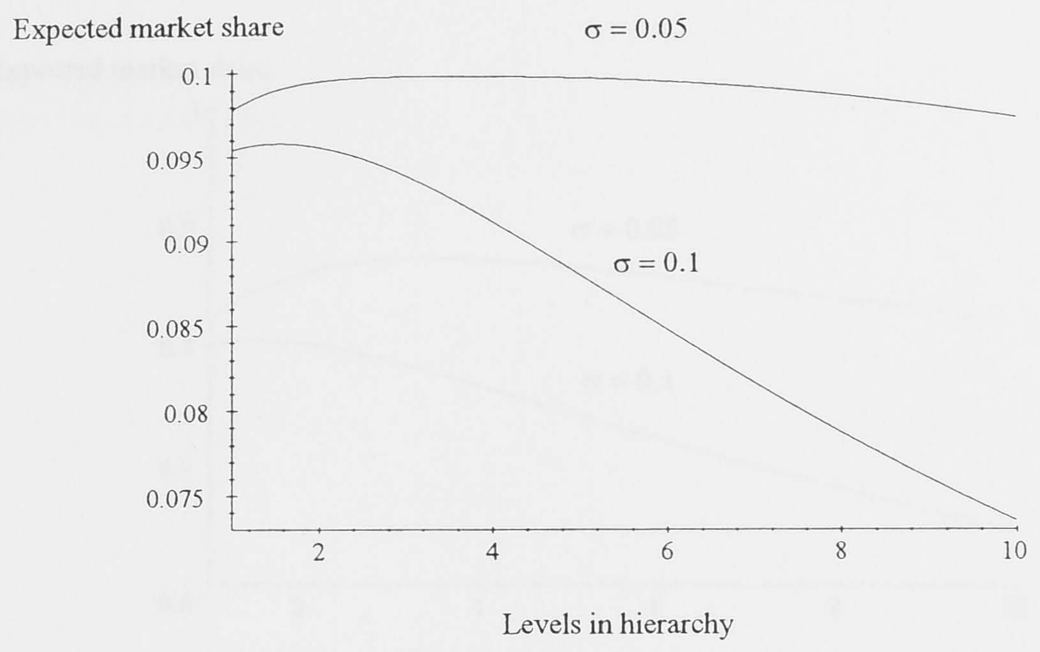


Figure 3.5: Expected market share for a specialized product ($w = 0.1$).

since the slopes of the two curves are more similar than in Figure 3.5. In part this is because the general product was less able to fulfill its potential market share even when shocks were small.

Instead of absolute performances we now consider relative performance.

$$\text{Expected percentage of potential market share} = \frac{\text{Expect market share } \times 100}{\text{Product width of appeal}}. \tag{3.14}$$

Recall, a product's width of appeal defines the maximum share of the market it could gain. Thus by this measure, if a firm which has $w = 0.1$ (the potential to gain 10% of the market) expects to get a market share of 0.1 then it has fulfilled 100% of its potential. This is an important real world consideration because it is the rate of return on investment, not the magnitude of profits, which is really important to investors. A specialized product should go hand in hand with plans for a lower level of production and hence less investment.

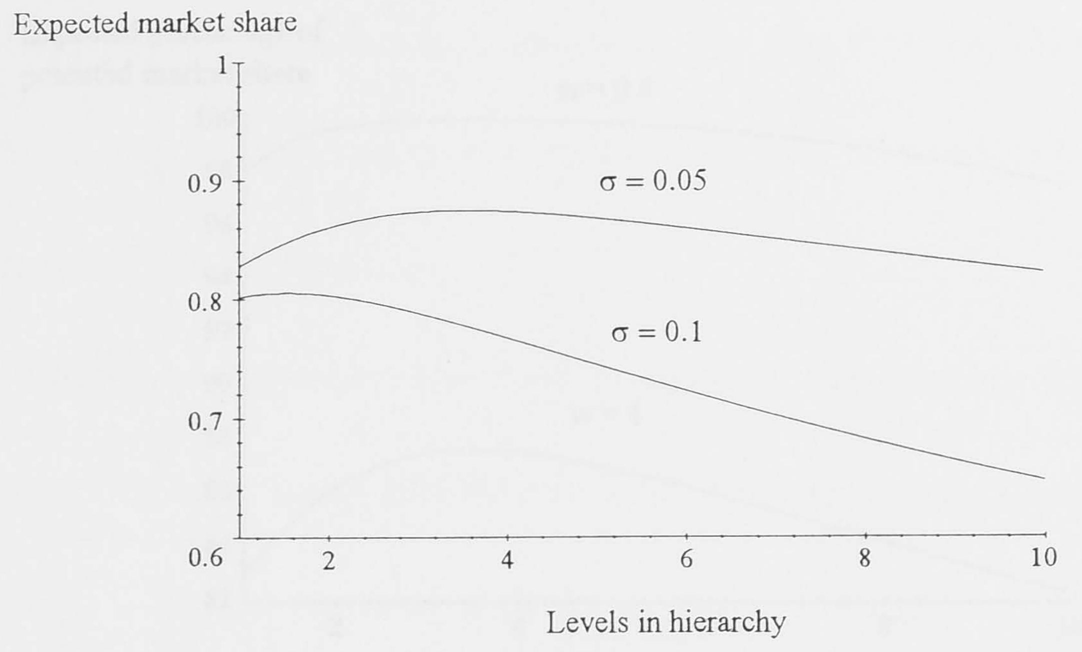


Figure 3.6: Expected market share for general product ($w = 1$).

Figure 3.7 shows that for a small shock, the specialized product is expected to fulfill more of its potential than the general product. Also the specialized product's ability to fulfill its potential is reasonably insensitive to a suboptimal choice of hierarchy. However as the magnitude of the shocks increases the relative performance of the two products converge, see Figure 3.8.

3.5 Conclusion

The first part of this chapter takes the standard characteristic space model of demand for differentiated products, and moves it to a dynamic stochastic setting. Consumer preferences change over time and are not directly observable, hence the decision of which good to produce becomes significant and difficult.

The firm solves the problem of which good to produce by using sample data

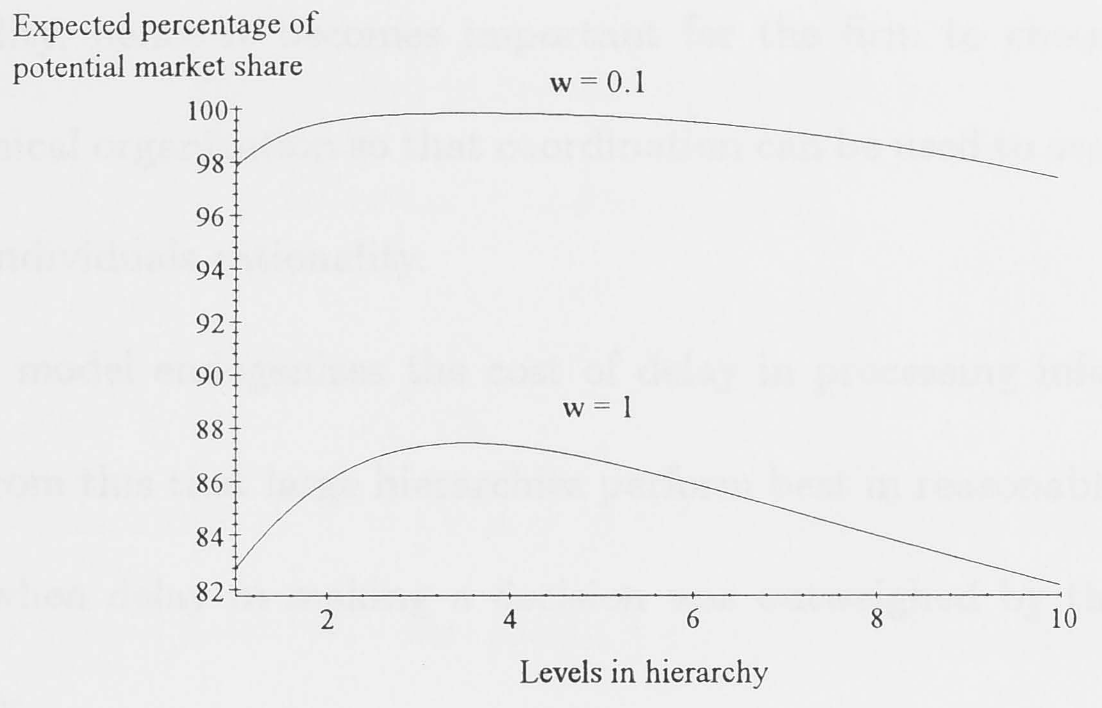


Figure 3.7: Expected percentage of potential market share for a specialized product and a general product when $\sigma = 0.05$.

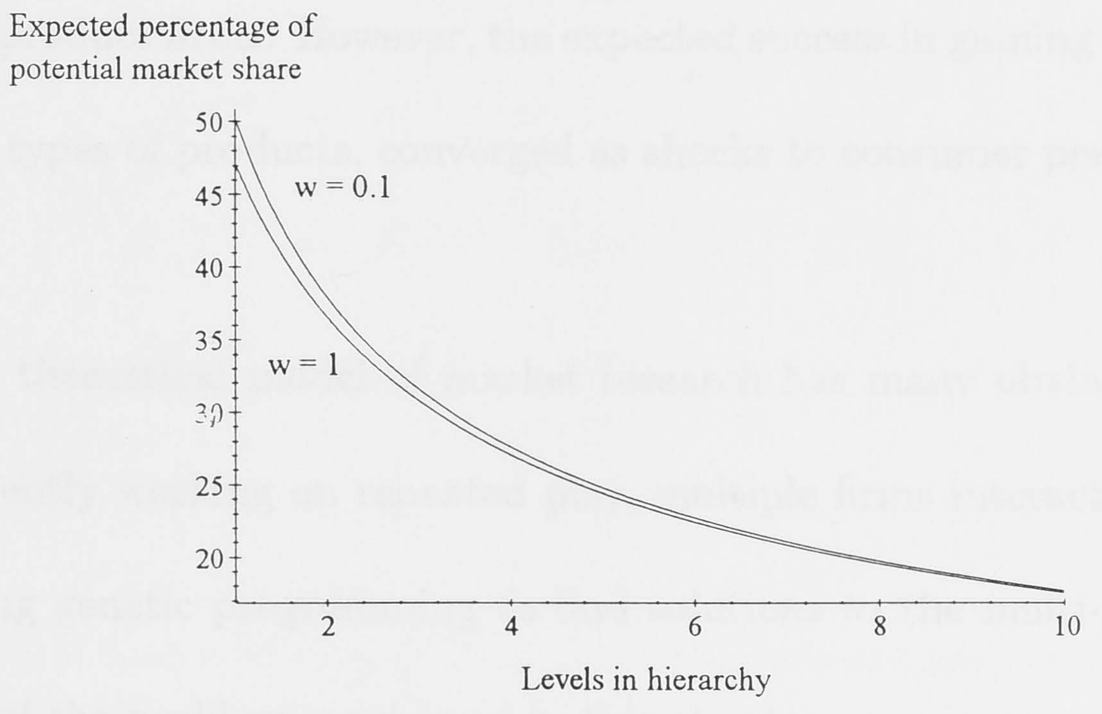


Figure 3.8: Expected percentage of potential market share for a specialized product and a general product when $\sigma = 0.5$.

from the market. However, the individuals using the sample data are of bounded rationality, hence it becomes important for the firm to choose some form of hierarchical organization so that coordination can be used to overcome the limits of the individuals rationality.

The model endogenizes the cost of delay in processing information. It followed from this that large hierarchies perform best in reasonably static environments when delay in making a decision was outweighed by the importance of accuracy.

It was also shown that the choice of hierarchy was less critical for a specialized product firm than a general product firm when consumer preferences experience only small shocks. Also counter intuitively specialized product firms were expected to be more successful at gaining their potential market share than general product firms. However, the expected success in gaining market share for the two types of products, converged as shocks to consumer preferences became large.

This theoretical model of market research has many obvious extensions: I am currently working on repeated play, multiple firms interacting strategically and using genetic programming to find solutions to the multi-period sampling version of the problem considered in this chapter.

3.6 Appendix: The First Four Moments of $g(z)$

We now turn to the problem of calculating the first four raw moments of $g(z)$. Recall $g(z)$ is the distribution of the expression

$$\theta + T - \overline{X}.$$

Where θ is a constant, T is the sum of δ Normally distributed shock terms and $\overline{X}$ is the sample mean of N independent draws from the uniform distribution on the interval $\left[\theta - \frac{1}{2}, \theta - \frac{1}{2}\right]$. The first four raw moments of these variables are as follows.

θ is a constant so $E[\theta^r] = \theta^r$ for $r = 1, \dots 4$.

By definition $T \tilde{} N(0, \delta\sigma^2)$, hence the first four moments of T are:

$$\begin{aligned}
E[T] &= 0 \\
E[T^2] &= E\left[(T - E[T])^2\right] = VAR[T] = \delta\sigma^2 \\
E[T^3] &= E\left[(T - E[T])^3\right] = 0 \\
E[T^4] &= E\left[(T - E[T])^4\right] = 3VAR[T]^2 = 3\delta^2\sigma^4
\end{aligned}$$

The first four moments of $\overline{X}$ are a little more involved to calculate because $\overline{X}$ is the average of N independent realizations of the random variable X which is uniformly distributed on $\left[\theta - \frac{1}{2}, \theta + \frac{1}{2}\right]$. The first step is to calculate the first four raw moments of X.

$$
\begin{aligned}
E[X] &= \theta \\
E\left[X^2\right] &= \int_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} x^2 dx = \left[\frac{x^3}{3}\right]_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} = \theta^2 + \frac{1}{12} \\
E\left[X^3\right] &= \left[\frac{x^4}{4}\right]_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} = \theta^3 + \frac{1}{4}\theta \\
E\left[X^4\right] &= \left[\frac{x^5}{5}\right]_{\theta-\frac{1}{2}}^{\theta+\frac{1}{2}} = \theta^4 + \frac{1}{2}\theta^2 + \frac{1}{80}.
\end{aligned}
$$

Using these moments we can now calculate the moments of $\overline{X}$. First the sample mean is an unbiased estimator of the population mean.

$$E\left[\overline{X}\right] = \theta.$$

The second moment requires expanding the sum:

$$
\begin{aligned}
E\left[\left(\overline{X}\right)^2\right] &= E\left[\left(\frac{1}{N}\sum X_i\right)^2\right] = \frac{1}{N^2}E\left[\left(\sum X_i\right)^2\right] \\
&= \frac{1}{N^2}E\left[\sum X_i^2 + \sum_{i \neq j} X_i X_j\right] \\
&= \frac{1}{N^2}\left[\binom{2}{2}\binom{N}{1}E\left[X^2\right] + \binom{2}{1}\binom{N}{2}E\left[X\right]^2\right]
\end{aligned}
$$

Substituting for the previously calculated moments

$$
\begin{aligned}
E\left[\left(\overline{X}\right)^2\right] &= \frac{1}{N^2}\left[N\left(\theta^2 + \frac{1}{12}\right) + N(N-1)\theta^2\right] \\
&= \theta^2 + \frac{1}{12N}
\end{aligned}
$$

We repeat the same process for the third moment.

$$
\begin{aligned}
N^3 E\left[\left(\overline{X}\right)^3\right] &= E\left[\sum X_i^3 + \sum_{i \neq j} X_i^2 X_j + \sum_{\substack{i \neq j \neq k \\ i \neq k}} X_i X_j X_k\right] \\
&= \left[N E\left[X^3\right] + \binom{3}{2,1}\begin{bmatrix} N \\ 2 \end{bmatrix} E\left[X^2\right] E[X] + \binom{3}{1,1,1}\binom{N}{3} E[X]^3\right] \\
&= \left[N E\left[X^3\right] + 3N(N-1) E\left[X^2\right] E[X] + N(N-1)(N-2) E[X]^3\right]
\end{aligned}
$$

Substituting and canceling

$$
\begin{aligned}
E\left[\left(\overline{X}\right)^3\right] &= \frac{1}{N^3}\left[N\left(\theta^3 + \frac{\theta}{4}\right) + 3N(N-1)\left(\theta^2 + \frac{1}{12}\right)\theta + N(N-1)(N-2)\theta^3\right] \\
&= \frac{\theta}{4N} + \theta^3
\end{aligned}
$$

Finally the fourth moment

$$
\begin{aligned}
N^4 E\left[\left(\overline{X}\right)^4\right] &= E\left[\sum X_i^4 + \sum X_i^3 X_j + \sum X_i^2 X_j^2 + \sum X_i^2 X_j X_k + \sum X_i X_j X_k X_m\right] \\
&= N E\left[X^4\right] + \binom{4}{3,1}\begin{bmatrix} N \\ 2 \end{bmatrix} E\left[X^3\right] E[X] + \binom{4}{2,2}\binom{N}{2} E\left[X^2\right]^2 \\
&\quad + \binom{4}{2,1,1}\begin{bmatrix} N \\ 1 \end{bmatrix}\binom{N-1}{2} E\left[X^2\right] E[X]^2 + \binom{4}{1,1,1,1}\binom{N}{4} E\left[X^4\right] \\
&= N\left(\theta^4 + \frac{1}{2}\theta^2 + \frac{1}{80}\right) + 4N(N-1)\left(\theta^3 + \frac{1}{4}\theta\right)\theta \\
&\quad + 6\frac{N(N-1)}{2}\left(\theta^2 + \frac{1}{12}\right)^2 + 12N\frac{(N-1)(N-2)}{2}\left(\theta^2 + \frac{1}{12}\right)\theta^2 \\
&\quad + 24\frac{N(N-1)(N-2)(N-3)}{24}\theta^4 \\
&= -\frac{N}{120} + \frac{1}{48}N^2 + \frac{1}{2}N^3\theta^2 + N^4\theta^4
\end{aligned}
$$

Therefore

$$
E\left[\left(\overline{X}\right)^4\right] = -\frac{1}{120N^3} + \frac{1}{48N^2} + \frac{1}{2N}\theta^2 + \theta^4.
$$

$$E\left[\theta + T - \overline{X}\right] = 0$$

$$\mu_1' = 0$$

$$\begin{aligned} E\left[\left(\theta + T - \overline{X}\right)^2\right] &= E\left[\left(\left(\theta + T - \overline{X}\right) - 0\right)^2\right] \\ &= VAR\left[\left(\theta + T - \overline{X}\right)\right] \\ &= VAR\left[\theta\right] + VAR\left[T\right] + VAR\left[\overline{X}\right] \\ &= \delta\sigma^2 + \frac{1}{12N} \end{aligned}$$

$$\mu_2' = \delta\sigma^2 + \frac{1}{12N}$$

$$E\left[\left(\theta + T - \overline{X}\right)^3\right] = E\left[\left(\theta + T - \overline{X}\right)^3\right]$$

$E\left[(\theta + T - S)^3\right]$

$= E\left[\theta^3 + 3\theta^2 T - 3\theta^2 S + 3\theta T^2 - 6\theta TS + 3\theta S^2 + T^3 - 3T^2 S + 3TS^2 - S^3\right]$

$= \theta^3 + 3\theta^2 E\left[T\right] - 3\theta^2 E\left[S\right] + 3\theta E\left[T^2\right] - 6\theta E\left[TS\right] + 3\theta E\left[S^2\right] + E\left[T^3\right] -$

$3E\left[T^2 S\right] + 3E\left[TS^2\right] - E\left[S^3\right]$

$= \theta^3 - 3\theta^2 E\left[S\right] + 3\theta E\left[T^2\right] + 3\theta E\left[S^2\right] + E\left[T^3\right] - 3E\left[T^2\right] E\left[S\right] - E\left[S^3\right]$

$= \theta^3 - 3\theta^3 + 3\theta\delta\sigma^2 + 3\theta\left(\theta^2 + \frac{1}{12N}\right) - 3\delta\sigma^2\theta - \left(\frac{\theta}{4N} + \theta^3\right)$

$= \theta^3 - 3\theta^3 + 3\theta\delta\sigma^2 + 3\theta^3 + \frac{1}{4}\frac{\theta}{N} - 3\delta\sigma^2\theta - \frac{1}{4}\frac{\theta}{N} - \theta^3$

$= 0$

$$\mu_3' = 0$$

That is there is no skewness.

$E\left[(\theta + T - S)^4\right]$

$= E\left[\begin{array}{c} 12\theta TS^2 - 12\theta^2 TS - 12\theta T^2 S + \theta^4 + 4\theta^3 T - 4\theta^3 S + 6\theta^2 T^2 \\ +6\theta^2 S^2 + 4\theta T^3 - 4\theta S^3 - 4T^3 S + 6T^2 S^2 - 4TS^3 + T^4 + S^4 \end{array}\right]$

$= 12\theta E\left[TS^2\right] - 12\theta^2 E\left[TS\right] - 12\theta E\left[T^2 S\right] + \theta^4 + 4\theta^3 E\left[T\right] - 4\theta^3 E\left[S\right] + 6\theta^2 E\left[T^2\right] +$

$6\theta^2 E\left[S^2\right] + 4\theta E\left[T^3\right] - 4\theta E\left[S^3\right] - 4E\left[T^3 S\right] + 6E\left[T^2 S^2\right] - 4E\left[TS^3\right] + E\left[T^4\right] + E\left[S^4\right]$

$= -12\theta E\left[T^2\right] E\left[S\right] + \theta^4 - 4\theta^3 E\left[S\right] + 6\theta^2 E\left[T^2\right] + 6\theta^2 E\left[S^2\right] - 4\theta E\left[S^3\right]$

$+6E\left[T^2\right] E\left[S^2\right] + E\left[T^4\right] + E\left[S^4\right]$

$= -12\theta^2\delta\sigma^2 + \theta^4 - 4\theta^4 + 6\theta^2\delta\sigma^2 + 6\theta^2\left(\theta^2 + \frac{1}{12N}\right) - \left(\frac{\theta}{4N} + \theta^3\right)$

$+6\delta\sigma^2\left(\theta^2 + \frac{1}{12N}\right) + 3\delta^2\sigma^4 - \frac{1}{120N^3} + \frac{1}{48N^2} + \frac{1}{2N}\theta^2 + \theta^4$

$= -12\theta^2\delta\sigma^2 + \theta^4 - 4\theta^4 + 6\theta^2\delta\sigma^2 + 6\theta^4 + \frac{1}{2}\frac{\theta^2}{N} - \frac{\theta}{4N} - \theta^3$

$+6\delta\sigma^2\theta^2 + \frac{1}{2}\delta\frac{\sigma^2}{N} + 3\delta^2\sigma^4 - \frac{1}{120N^3} + \frac{1}{48N^2} + \frac{1}{2N}\theta^2 + \theta^4$

$= 4\theta^4 + \frac{\theta^2}{N} - \frac{1}{4}\frac{\theta}{N} - \theta^3 + \frac{1}{2}\delta\frac{\sigma^2}{N} + 3\delta^2\sigma^4 - \frac{1}{120N^3} + \frac{1}{48N^2}$

$$\mu_4' = 4\theta^4 + \frac{\theta^2}{N} - \frac{1}{4}\frac{\theta}{N} - \theta^3 + \frac{1}{2}\delta\frac{\sigma^2}{N} + 3\delta^2\sigma^4 - \frac{1}{120N^3} + \frac{1}{48N^2}$$

Chapter 4

A Flexible Price, Monopoly Location Problem with Hierarchies

4.1 Introduction

In the previous chapter a firm with an internal decision making structure was made the protagonist in a problem from industrial organization. In this chapter the analysis is extended beyond the simple case considered in the previous chapter.

The situation is again that of a firm launching a product into a market where consumer preferences change over time and are not directly observable at the population level by the firm. The firm will have to use signals on market conditions in order to forecast the state of preferences at the launch date (which will be referred to as the location of the market).

As before, using a forecast of market conditions will generally produce an error between the chosen location and the profit maximizing location. We will assume that once chosen, the firm's location remains fixed but we now allow the firm to adjust its price to maximize profits, given its location.

The second extension in this chapter is to impose no constraint on the class of hierarchies considered. In fact using a result from Radner(1993), the optimal hierarchies for the modeled decision problem will be analyzed.

Owing to unmodeled institutional constraints, actual firms may not use an optimal hierarchy. Similarly a firm may not be able to adjust instantaneously the price of its product after launch. In these circumstances, the analysis in this chapter provides an upper bound on a firm's performance in this type of market. Nevertheless the assumptions of this chapter do apply to real world situations as the following example shows.

Computer software is a product who's price can be adjusted after its launch, but who's characteristics remain fixed (in the short term at least). As the following example, from Carroll(1993, p86), of the IBM product TopView shows, optimal pricing after launch cannot undo the problem of having misread the market due to inappropriate organizational structure.

"...the idea would have let people use more than one program at

> once by allowing them to divide their screens into various windows. The idea would also have reduced complexity by giving them menus of commands to choose from,... As the months went on, however, IBM began to botch its latest software project for all the usual reasons - it put too many people on the project, the work took too long, the software operated too slowly, and it turned out that customers wanted something much glitzier than IBM provided."

TopView was a precursor of Microsoft's extremely successful Windows software, and in that respect was based on a sound premise. Although offering more advanced features than DOS and launched without a great deal of competition (Windows was not available until 1985) no pricing strategy could overcome IBM's problem of failing to match the direction the market would take:

> "(TopView) was dubbed TopHeavy by customers and became one of the biggest flops in the history of IBM's PC business. IBM wound up giving away most of the copies of TopView that it produced after its introduction in 1984."

This chapter attempts to address two main questions motivated by this example. What kind of management structure should a firm in this kind of market adopt in order to try and maximize profits? What effects do differing management structures have on performance?

Bounded rationality, or if you prefer, the limited ability of humans to deal with information, is the key to understanding management structures. The intuitive explanation is that one individual cannot organize everything because of his/her finite mental capacity. Hence the task of running a firm is best split amongst a number of people.

One simple version of this approach is to assume that performing management tasks takes time and that each individual can only do one thing at a time. If it is important to perform management tasks in a timely manner, then there will be benefits from splitting tasks amongst a number of people so that different aspects of the same problem can be worked on at the same time.[1]

Section 4.3 shows how the results on optimal hierarchical structure from Radner(1993) can be applied to the real time decision problem considered here. As in the previous chapter the decision to be made will be the product location based on the forecast of market conditions, calculated from raw data on individual consumer preferences. The optimal choice of organizational structure is the one which maximizes expected profit. Section 4.2.1 defines the market conditions the firm faces and derives the profit maximizing price (and associated level of profit), given a location.

The type of organizational structure determines the distribution of the error in the forecast of market conditions. In section 4.4 the distribution of the forecast error and the profit function (as a function of forecast error) are combined to

[1]This is the key intuition underlying Radner(1993)

give an expression for expected profit as a function of organizational structure and various parameters describing market conditions. A solution method for this optimization problem is also discussed in section 4.4 and the results are presented in section 4.5.

4.2 Location then Price Model

The particular model which we will use has two stages: a monopolist chooses location and then price. Using its management structure, the firm decides where to locate its product at the launch date. Upon fixing the location of its product, market location is revealed to the firm, which sets its price accordingly to maximize profits.

In this section, we describe the market which the firm faces and then calculate the profit maximizing price response for a given location.

4.2.1 Market Conditions and the Optimal Price

Following the standard specification, it is assumed that products have a characteristic that consumers have preferences over, and that this characteristic can be represented by a real number x. Each consumer has a unique, most preferred value of the characteristic, called their ideal point, also represented by a real number c. The greater the difference between a consumer's ideal point and a particular product, the lower the utility that product provides the consumer.

Both products and ideal points must be in the characteristic space, which is assumed here to be the whole real line.

To be specific, it is assumed that the indirect utility function or valuation function $\widehat{U}(\cdot)$ for product x by consumer c has the following linear form in distance

$$\widehat{U}(x,c) = \widehat{a} - \widehat{p} - \widehat{\kappa}\,|x-c|\,.$$

Where $\widehat{a}$ is the common valuation of the product, $\widehat{p}$ is the price and $\widehat{\kappa}$ the rate at which utility decreases with distance. The coefficient $\widehat{\kappa}$ is typically referred to as the transport cost. It is assumed that the price, valuation, transport cost and the functional form are the same for each consumer. The only difference between consumers is their location.

Without loss of generality the valuation function can be renormalized so that the common valuation is 1. This implies that the new valuation function will have the form

$$U(x,c) = \frac{1}{\widehat{a}}\widehat{U}(x,c) = 1 - p - \kappa\,|x-c|\,.$$

Where $p = \frac{\widehat{p}}{\widehat{a}}$ and $\kappa = \frac{\widehat{\kappa}}{\widehat{a}}$. We assume there is some outside option giving utility 0, and that consumers have sufficient income to purchase for $p \in [0,1]$.[2] It follows that consumer c will only purchase x if $U(x,c) \geq 0$.

[2]Under the renormalization the maximum valuation a consumer can have for a product is 1. Thus the firm will not charge a price in excess of 1 because if it did no consumers would choose to purchase its product.

The width of appeal of a product, $w(x,p,\kappa)$, is the section of the characteristic space in which consumers would have to be located in order to purchase the product, that is, for a given x, the set of c for which for $U(x,c) \geq 0$. Under this specifications a product's width of appeal is given by

$$w(x,p,\kappa) = \left[x - \frac{1-p}{\kappa}, x + \frac{1-p}{\kappa}\right].$$

The quantity demanded of the product will in general be less than w because only those consumers whose actual locations are within w will purchase x. We assume that consumers are uniformly distributed on the interval $\left[\mu_t - \frac{1}{2}, \mu_t + \frac{1}{2}\right]$.[3] This population distribution is referred to as the market. We assume that the whole population of consumers gives rise to at most one unit of demand[4]. Thus, the actual quantity demanded is determined by the intersection of w and the population of consumers $\left[\mu_t - \frac{1}{2}, \mu_t + \frac{1}{2}\right]$. For this reason the quantity demanded will be referred to as the population share, denoted s. The level of s is

$$s(x,p,\kappa,\mu_t) = m\left[w(x,p,\kappa) \cap \left[\mu_t - \frac{1}{2}, \mu_t + \frac{1}{2}\right]\right].^{5}$$

Assume that $\kappa \geqslant 2$. This is the more economically interesting case, where for any price $p > 0$, even if the monopolist is located at the centre of the market not all consumers will choose to purchase from it.

[3]The process which determines μ_t is defined later.

[4]Again this is just a matter of scaling.

[5]$m[]$ is the Lebesgue measure

Production incurs a fixed cost, F, and zero marginal cost. Let π denote profit and let r denote revenue. The monopolist's profit maximization problem is

$$\begin{aligned} \max_{p,x} E\left[\pi\left(x, p, \mu_t, \kappa\right)\right] &= E\left[r\left(x, p, \mu_t, \kappa\right) - F\right] \\ &= E\left[r\left(x, p, \mu_t, \kappa\right)\right] - F. \end{aligned}$$

Hence profits will be maximized if expected revenue, $E\left[r\left(x, p, \mu_t, \kappa\right)\right]$ is maximized.

We first consider the profit maximizing price, p^*, for each location x and the associated revenue r^*, within a period (so that μ_t is fixed). Only the position of x relative to the market (μ_t) is important, so without loss of generality we consider $z = (x - \mu_t)$, which is referred to as the location error. A number cases, depending on the relative magnitudes of z and κ, need to be considered in the calculation of p^*. The details are in Appendix A of this chapter.

$$p^*(z, \kappa) = \begin{cases} \frac{1}{2} & \text{if } \; 0 \leq |z| \leq \frac{1}{2} - \frac{1}{2\kappa} \\ \kappa\left(|z| + \frac{1}{\kappa} - \frac{1}{2}\right) & \text{if } \; \frac{1}{2} - \frac{1}{2\kappa} \leq |z| \leq \frac{1}{2} - \frac{1}{3\kappa} \\ \frac{\kappa}{2}\left(\frac{1}{2} - |z| + \frac{1}{\kappa}\right) & \text{if } \; \frac{1}{2} - \frac{1}{3\kappa} \leq |z| \leq \frac{1}{2} + \frac{1}{\kappa} \\ 0 & \text{if } \; |z| \geq \frac{1}{2} + \frac{1}{\kappa} \end{cases}.$$

Substituting p^* into $r\left(p, z, \kappa\right)$ gives $r^*\left(z, \kappa\right)$ the profit maximizing revenue:

$$r^*(z, \kappa) = \begin{cases} \frac{1}{2\kappa} & \text{if } \; 0 \leq |z| \leq \frac{1}{2} - \frac{1}{2\kappa} \\ \left(2 + 2\kappa|z| - \kappa\right)\left(\frac{1}{2} - |z|\right) & \text{if } \; \frac{1}{2} - \frac{1}{2\kappa} \leq |z| \leq \frac{1}{2} - \frac{1}{3\kappa} \\ \frac{1}{16\kappa}\left(-\kappa - 2 + 2|z|\kappa\right)^2 & \text{if } \; \frac{1}{2} - \frac{1}{3\kappa} \leq |z| \leq \frac{1}{2} + \frac{1}{\kappa} \\ 0 & \text{if } \; |z| \geq \frac{1}{2} + \frac{1}{\kappa} \end{cases}.$$

Once the probability density $f\left(z\right)$ of z has been determined, the expected revenue is calculated by

$$E\left[r^*\right] = \int_{-\left(\frac{1}{2}+\frac{1}{\kappa}\right)}^{\frac{1}{2}+\frac{1}{\kappa}} r^*(z) f\left(z\right) dz.$$

In order to determine $f(z)$ we must first describe the process by which a location x is chosen. We turn to that problem now.

4.3 Decision Making and Organizational Structure

We now consider the process by which a particular organizational structure determines a forecast and hence decides where to locate the product. In section 4.4 we examine how the different organizational structures affect expected revenue. There are two parts to the forecasting process that we model: the stochastic process which generates the observations (considered next) and the organizational structure which processes the observations into a forecast (described in section 4.3.2).

4.3.1 The Real Time Decision Problem

Consumer preferences change over time, that is, the distribution of ideal points changes. It is assumed that only one parameter μ_t changes. Assume that μ_t follows a known data generating process, namely

$$\mu_t = \mu_{t-1} + \varepsilon_t$$

with $\varepsilon_t \sim N(0, \sigma^2)$.

The assumption that the firm only has to learn about one parameter is purely for expositional ease, and is typical of the literature on firms learning about demand conditions, for example Harrington (1995). With little modification the following could describe a firm using kernel estimation in order to build up an estimate of the distribution of consumers. This would add little insight and a lot more detail because, for example, the derivation of r^* in the previous section would have to be based on the prior distribution of all possible population distributions.

The firm knows the form of the process for μ_t but does not observe the actual values of μ_t. What the firm can observe is the ideal points of those particular consumers whom it surveys during a particular time period. The set $C_t = \{c_1, ..., c_N\}$ consists of the ideal points of the N consumers surveyed in period t. This is what we refer to as market research: ask someone what they like and they tell you what their ideal point is[6].

Two assumptions are made here about the market research: surveys of consumer ideal points are the only relevant information available and only one such survey is undertaken. The second assumption does have some justification because there is a large fixed cost incurred when a survey is conducted. Also products such as consumer software are fairly easy to duplicate, thus a firm will not want to make test copies too widely available before the final launch of the product. Relaxing the assumption of a single sample is quite hard technically, the process of updating the estimate of μ_t on the basis of new information is

[6]It is assumed that there is no strategic misrepresentation.

solved by the Kalman Filter, and Meagher and Miron (1997) show that a version of the Kalman Filter is amenable to parallel computation. The remaining problem is to characterize the optimal organizational structure for the parallel computation of the Kalman Filter, this is the subject of on going research. For the moment we consider only the single sample case, which we will show is sufficiently rich as to produce a number of interesting results.

The firm's decision making problem is to decide on a sample size and time period in which to survey consumers, an organizational structure with which to deal with all this information and a location based on the processed survey information. The decisions on the quantity of information, the period for information collection and the organizational structure will be inter-related because of the assumption that dealing with information takes time. The exact relationship is considered next.

4.3.2 Individuals and Organizational Design

The limited capabilities of a firm's employees are explicitly modelled by describing each individual as a processor and the organization as a programmed network. We will briefly restate the appropriate definitions.

A processor has an *inbox* and a *register* (which together comprise its *memory*). In one period of time, a processor can perform all or some of the following operations in the following order:

1. Read a number from the sample or its inbox and store it in its register, overwriting the previous contents of the register.

2. Take one number, y, from its inbox, calculate a binary associative function of y and the contents of its register and store the result in its register.

3. Write the contents of its register to the inbox of any processor to which it is connected (this is called *output*).

Together processors and connections form a *network*. A *program* specifies which operations each processor performs in each period.

There is only one decision to be made by the network: which product to make. It is shown in Radner(1993) that for this type of decision, being made by a network of processors, it is sufficient to consider only hierarchies. To describe the time taken to make the decision we use the concept of delay. Delay, D, is the amount of time a hierarchy takes from the arrival of the first piece information to be used until the decision is produced.

Let one sample consumer ideal point be one unit of information. There are N units of information to be considered in C_t which all arrive at the same time. Let P be the number of employees (also referred to as managers) who are involved in processing the N pieces of information. It is shown in Radner(1993) that $\widehat{D}$ the minimum delay for a given N and P is

$$\widehat{D}(N,P) = \left\lfloor \frac{N}{P} \right\rfloor + \lceil \log_2 (P + N \bmod P) \rceil . \qquad (4.1)$$

It follows that for a fixed N, the lowest possible value of $\widehat{D}(N, P)$ is achieved with $P = \left\lceil \frac{N}{2} \right\rceil$. The hierarchy which achieves this minimum delay for a fixed N is called the delay-is-everything hierarchy.

The proof in Radner(1993) of these results is somewhat involved. However, Radner(1993) provides an algorithm for converting uniform binary hierarchies (which were considered in the previous chapter) into hierarchies with minimum delay for a given sample size and a given number of employees. The details of the algorithm will not be repeated here, but the intuition for the method gives insight into the character of optimal hierarchies.

In a regular binary hierarchy the supervisors of the level which is currently processing are idle. Time could be saved if some of those in the current level were eliminated and their work passed to their supervisors. Time is saved in this way because there is an implicit cost in rereading information every time a report is passed upwards. This can be somewhat alleviated by eliminating some unnecessary reports (and the individuals who authored them). Repeating this procedure until no more improvements can be made gives an optimal solution.

For a given N, there are $\left\lceil \frac{N}{2} \right\rceil$ managers in the delay-is-everything hierarchy, hence the choice of hierarchy can again be described by one parameter, N. These hierarchies have skip level reporting, do not have a regular shape and are not as readily interpreted in terms of levels as the completely uniform hierarchies in the previous chapter. However, conclusions about the size of the management structure can still be drawn in terms of the number of managers.

The model should not be thought of as a literal attempt to describe the process of market research, rather the model tries to capture the interactions of information, organizational structure and product market decisions.

The relevant binary associative function will turn out to be a simple linear function of real numbers in this model. The adding of numbers should be considered as an analogy of the processing of information and the production of reports, recommendations and decisions that really go on inside the management structure of an organization. It might seem a trivial task to collate sample information, but the information which flows within this programmed network should also be thought of as the decisions and recommendations that are based on raw information.

The design of a real product is much more complex than choosing a number for its characteristics. For example, IBM learned that people wanted an operating system that was simpler to use than DOS. Converting that information into a product strategy is an enormously complex task. Without even producing the actual product IBM would have to make decisions on which platforms it would run under, what kind of interface (words or graphics), whether business users all want the same thing, what the demand conditions are in markets broken down by user type and geographic location and so on. All these things would have to be decided on from market research data. This is the process being described by the processing of consumer ideal points.

Since the process which generates μ_t is non stationary there is no long run expected value for μ_t, hence, a firm must use some market research to condition its product location otherwise it will have zero expected profit.

As noted in the previous chapter, there is not as yet a method for endogenizing the level of wages in the information processing model. Thus wages is treated as a parameter. An arbitrary choice of wages offers little insight so we assume here that the level of wages is sufficiently low as to not affect the choice of organizational structure. This implies that the firm will only be choosing from amongst the set of delay-is-everything hierarchies.

The minimum delay possible for each size N, is given by $1 + \lceil \log_2 N \rceil$. Since there is no cost for extra information, only those N which are powers of 2 need be considered.

4.4 Maximizing Expected Profits

We have now specified the process by which a location x is calculated, all that remains is to determine the actual rule used to determine the location and the distribution of the location error it will produce.

Assume that the survey was taken in period t. The sample $C_t = \{c_1, \ldots c_N\}$ that a firm receives is the set of ideal points for the N consumers who were surveyed. By calculating the sample mean $\overline{C}_t$, the firm has an unbiased estimate of μ_t, which was the mean of the distribution from which the sample came. Since this sample is the only information on which to condition expectations it follows that

$$E[\mu_t|C_t] = E[\mu_{t+D}|C_t] = \overline{C}_t, \qquad \forall D > 0. \tag{4.2}$$

Where D is the number of periods of delay caused by calculating $\overline{C}_t$.

So the decision rule for the firm will be of the form $x = \overline{C}_t + \Sigma$ where Σ is some strategic move the firm wants to make away from its estimate of μ_{t+D}. It now remains to show that $\Sigma = 0$ is the optimal choice.

Now the market shifts according to a random walk with Normally distributed shocks, so μ_{t+D} can be rewritten as

$$\begin{aligned} \mu_{t+D} &= \mu_t + \sum_{i=1}^{D} \varepsilon_{t+i} && (4.3) \\ &= \mu_t + T. && (4.4) \end{aligned}$$

Where $T = \sum_{i=1}^{D} \varepsilon_{t+i}$, is the sum of the shock terms. The shock terms are independent and identically distributed, $\varepsilon_t \sim N(0, \sigma^2)$, hence $T \sim N(0, D\sigma^2)$.

r^* is a function of $z = (\mu_t + T - x)$, the distance between the centre of the market and the actual location in period $t + D$. Also r^* is symmetric about $z = 0$, non increasing in general and decreasing for sufficiently large z.

T has infinite support therefore z also has infinite support. It was shown in the appendix to the proceeding chapter that the distribution $g(z)$ (for $z = (\mu_t + T - \overline{C})$) is symmetric. Therefore choosing $\Sigma \neq 0$ would make lower revenue z's more likely, while making higher revenue z's less likely. Thus, it is optimal to choose $\Sigma = 0$.

It is thus possible to calculate the moments of the forecast error. For example

$E[z] = E[\mu_t(t)+T-\overline{C}] = 0$ and $Var[z] = Var[\mu_t(t)+T-\overline{C}] = Var[T]+Var[\overline{C}]$.

$$Var[T] + Var[\overline{C}] = (1 + \log_2 N)\, \sigma^2 + \frac{1}{12N}. \tag{4.5}$$

The distribution of $g(z)$ is not known, so we could, as in the proceeding chapter, approximate $g(z)$ by its asymptotic distribution, which is Normal. This would involve matching the first two moments of a Normal distribution with the mean and variance calculated above. Instead we construct the expected revenue curve by Monte Carlo simulation. The details of the simulation procedure are presented in the appendix 4.8.

4.5 Results

We now present the results of the Monte Carlo simulations. The underlying distributions are all known, so for specific values of the parameters the expected revenue function can be generated by simulation.

The results from some interesting simulations are shown in Figures 4.1-4.3. Each Monte Carlo is based on 10000 repetitions so one can be quite confident that these curves, and the relationships between them, are accurate.

Figures 4.1 and 4.2 both show that increasing σ for a fixed κ leads to a lower level of expected revenue. This occurs because increasing σ increases the variance of the forecast error. For a fixed κ the function r^* is fixed and non-increasing in the magnitude of the forecast error. Therefore increasing the variance of the

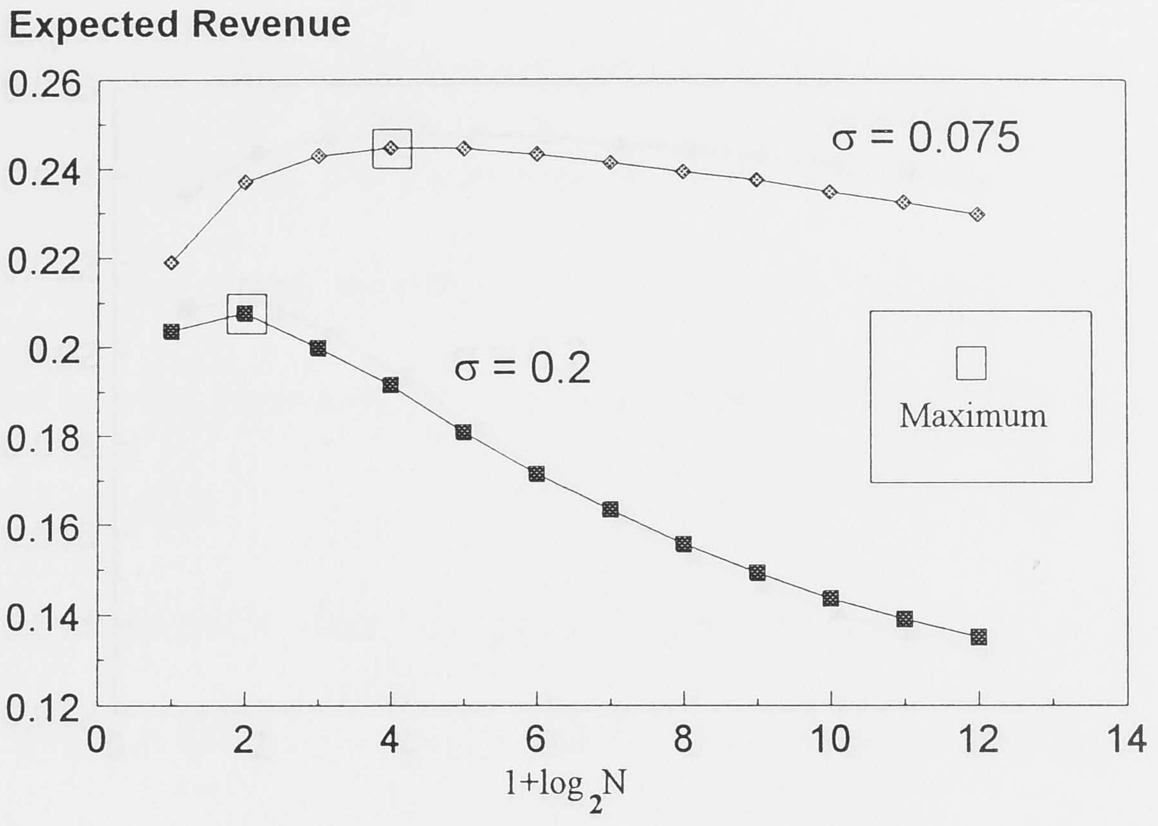


Figure 4.1: Expected revenue for $\kappa = 2$

forecast error causes high revenue outcome to become less likely while low revenue outcomes become more likely, thus decreasing expected revenue.

These two figures also show that the optimal organizational size becomes smaller as σ increases. Increasing σ means that every period of delay will cause even more uncertainty about the future (greater dispersion of $f(z)$). However the benefit of information (N) has remained the same, thus as σ increases it becomes optimal for a firm to use less information and to act faster, and hence have a smaller organization.

Comparing Figure 4.1 with Figure 4.2 we see that for a fixed σ, changing the value of κ has no effect on the optimal organization size. When $\sigma = 0.2$ the optimal organization for both $\kappa = 2$ and $\kappa = 20$ is one person making a decision

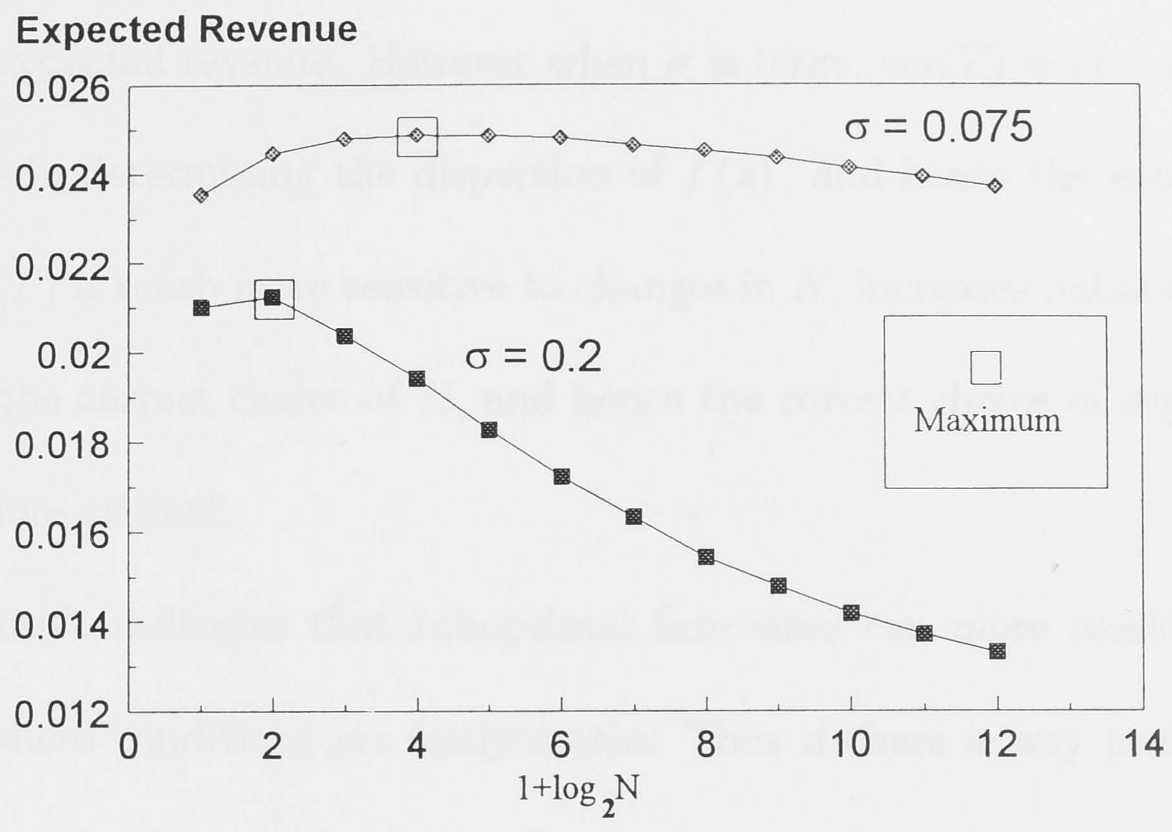


Figure 4.2: Expect revenue for $\kappa = 20$

based on two pieces of information. When $\sigma = 0.075$, the optimal organization for the two levels of κ is four people using eight sample points to make a decision. These results indicate that in finding the best organization for forecasting, the firm does not need to consider the type of product it is producing because, intuitively, N appears in $f(z, \sigma, N)$ but not in r^*.[7]

In both Figures 4.1 and 4.2 the expected revenue curve is flatter for the lower value of σ than for the higher value. When σ is low, most of the dispersion in $f(z)$ comes from the difference between μ_t and $\overline{C}_t$. The variance of this difference is $\frac{1}{12N}$ which changes by only a small amount as N varies for sufficiently large N.

[7]A formal proof of this result would require consideration of tail behavior of both r^* and f, which is intractable because the form of f is unknown.

Thus when σ is small, suboptimal choice of organizational structure has little effect on expected revenue. However when σ is large, $\text{var}(T) = (1 + \log_2 N)\,\sigma^2$, dominates in determining the dispersion of $f(z)$, and hence the expected revenue. $\text{var}(T)$ is much more sensitive to changes in N, increases unbounded in N, therefore the correct choice of N, and hence the correct choice of organization, become more critical.

This result indicates that suboptimal firm sizes can more readily exist in markets where conditions are fairly static. Thus if there is any preference for management largess, in the form of large companies, it is more likely to be observed in stable markets. Fast changing markets penalize firms much more for suboptimal organizational choices.

Expected revenue is increasing in κ by definition. A larger κ means a larger w so a product must be able to generate more revenue. What about relative performance? Relative performance can be compared by considering scaled revenue, r_s. Where

$$r_s(p, x, \kappa) = r(p, x, \kappa) \times 2\kappa.$$

Figure 4.3 shows scaled expected revenue for two types of products: a general product with a low transport cost of $\kappa = 2$ and a specialized or niche product with a transport cost of $\kappa = 20$. At the same price the width of appeal for the mass appeal product is 10 times wider than for the specialized product.

We see that scaled revenue for the specialized product is higher than for

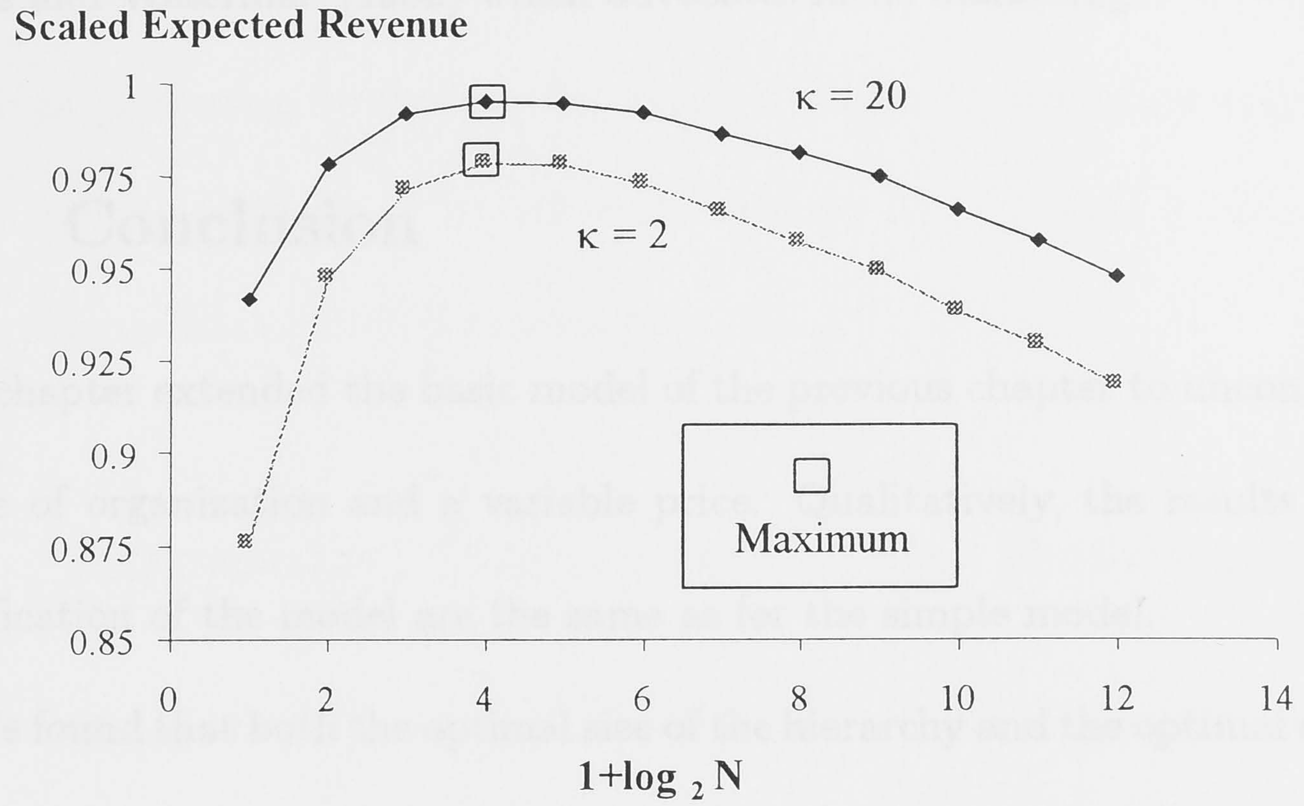


Figure 4.3: Scaled expected revenue for $\sigma = 0.075$

the mass appeal product. For the optimal organizational choice, the specialized product has an expected revenue which is almost 100% its maximum revenue. Expected scaled revenue for the general product shows that its expected revenue is 98% of its maximum revenue. Thus we observe that specialized products do better than general products in terms of fulfilling their revenue potential. However the difference for this parameterization of the flexible price situation is less dramatic than was seen for the fixed price simulations of the previous chapter. The result appears to hold for other parameter values in the variable price model, but has not been examined in detail.

If the fixed cost F is directly proportional to the transport cost κ then the specialized product would show a better rate of return than the mass appeal product. This corresponds to the anecdotal marketing evidence presented in

Peters and Waterman (1982) which advocates niche marketing.

4.6 Conclusion

This chapter extended the basic model of the previous chapter to unconstrained choice of organization and a variable price. Qualitatively, the results for this specification of the model are the same as for the simple model.

We found that both the optimal size of the hierarchy and the optimal quantity of information arise endogenously when both the information processing abilities of managers and the economic problem are specified. The effects of market conditions on profitability and organization choice were also consider. It was shown that increasing the variance of shocks to consumer preferences decreases the optimal firm size and expected profits. Changing the transport cost of a product had no impact on the choice of organization. However, narrow appeal products (high transport cost) had an expected revenue which was a higher proportion of their maximum revenue that did mass appeal (low transport cost) products. This suggested that if the fixed cost is proportional to the transport cost then specialized products will have a better rate of return. These results on the performance of different types of products, specialized or general, are only preliminary but would be an interesting area for future research.

There are a number of obvious technical extensions to the model, such as multiple period sampling instead of single period sampling, preliminary work on this subject can be found in Meagher and Miron (1997). More generally, the approach of making hierarchies the decision makers in industrial organization games seems promising. I am currently working on a duopoly location model and an oligopolistic product innovation model.

4.7 Appendix: Derivation of the Revenue Function

Recall $z = x - \mu_t$. Due to symmetry, we can consider, without loss of generality, only the cases where $x > \mu_t$. There are four such cases, depending on the magnitude of z relative to p and κ, which we will now considered in turn.

First, if $z \geq \frac{1}{2} + \frac{1}{\kappa}$ then all of $w(z, p)$will lie outside the market, for all $p \geq 0$. In this case, revenue will be zero for all $p \geq 0$. We will refer to this as case 0.

The other three cases are less straight forward. Case 1 is when the price, p_1, is such that $w(p_1, z)$ lies within the market $\left[-\frac{1}{2}, \frac{1}{2}\right]$. If this is the case, then the population share will be given by

$$\begin{aligned} s(p_1, z, \kappa) &= z + \frac{(1 - p_1)}{\kappa} - \left(z - \frac{(1 - p_1)}{\kappa}\right) \\ &= \frac{2(1 - p_1)}{\kappa}. \end{aligned}$$

This situation is depicted in Figure 4.4. The revenue for case 1, $r_1(p, z, \kappa)$, is given by

$$r_1(p, z, \kappa) = p_1 \frac{2(1 - p_1)}{\kappa}.$$

The first order condition for the revenue maximizing price, p_1^* is

$$\frac{2}{\kappa} - \frac{4p_1^*}{\kappa} = 0 \Rightarrow p_1^* = \frac{1}{2}.$$

Hence the maximum revenue under case 1, r_1^* will be

$$= \frac{1}{2}\frac{2\left(1-\frac{1}{2}\right)}{\kappa} = \frac{1}{2\kappa}.$$

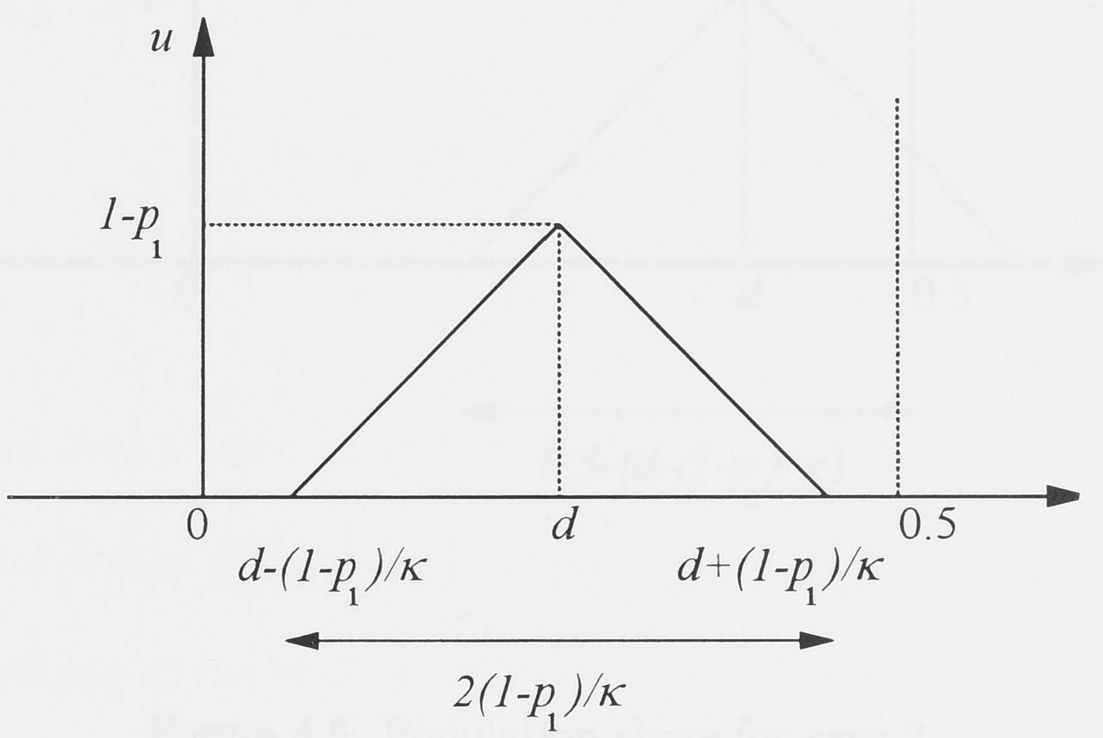


Figure 4.4: Population share for Case 1.

We now turn to finding the range of z for which p_1^* is the revenue maximizing price and $r^*{}_1$ the maximum revenue. By assumption case 1 is true only if $w(z, p_1)$ is within $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Since z is non negative, this will only be true if $z + \frac{1-p_1}{\kappa} \leq \frac{1}{2}$. However we know that in order to maximize revenue under case 1 the firm will set $p_1 = p_1^* = \frac{1}{2}$. Thus the condition

$$z + \frac{1-p_1^*}{\kappa} = z + \frac{1}{2\kappa} \leq \frac{1}{2}$$

must hold. Rearranging gives the region over which p_1^* does indeed maximize revenue at r_1^* :

$$0 \leq z \leq \frac{1}{2} - \frac{1}{2\kappa}.$$

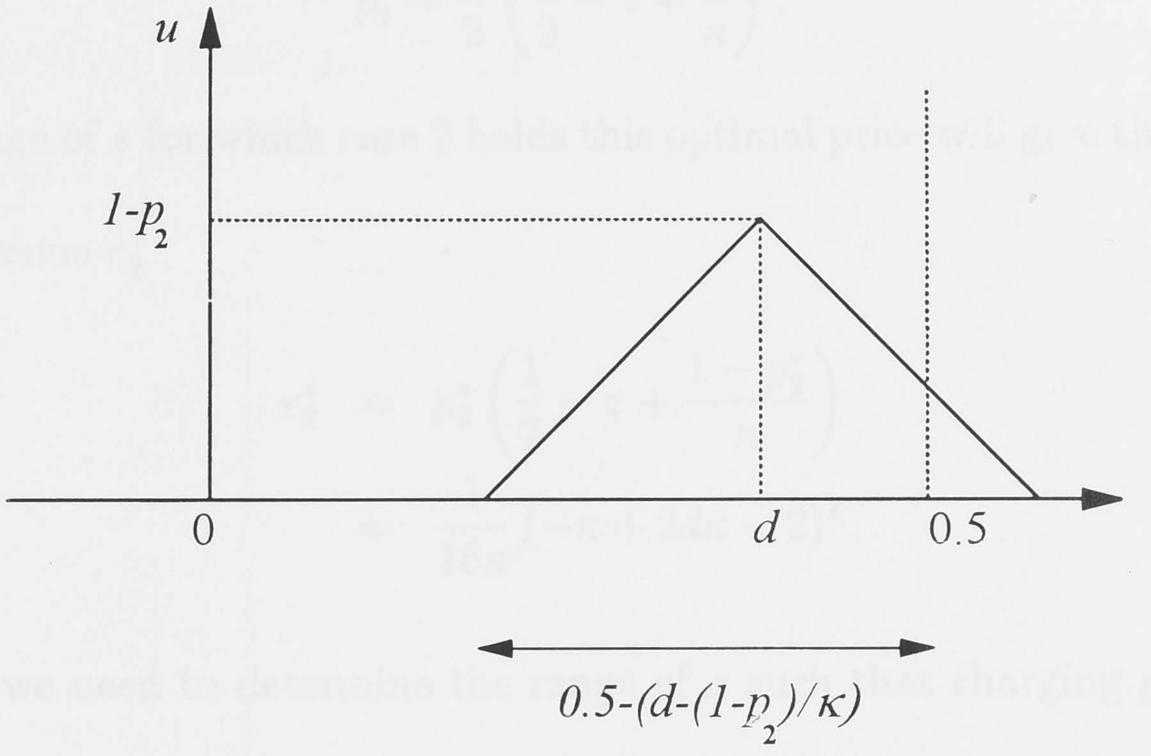


Figure 4.5: Population share for case 2.

Under case 2, the price p_2 is such that part of $w(z, p_2)$ lies outside the market. This situation is shown in Figure 4.5. Consumers who will purchase the product lie between $z - \frac{1-p_2}{\kappa}$ and the edge of the market/population at $\frac{1}{2}$. Thus the population share is

$$s(z, p_2, \kappa) = \frac{1}{2} - \left(z - \frac{1-p}{\kappa}\right)$$

and hence the revenue under case 2, $r_2\left(z, p_2, \kappa\right)$ is

$$r_2 = p_2\left(\frac{1}{2} - z + \frac{1-p_2}{\kappa}\right).$$

The first order conditions for revenue maximization are

$$\frac{1}{2} - z + \frac{1 - 2p_2^*}{\kappa} = 0.$$

Rearranging gives

$$p_2^* = \frac{\kappa}{2}\left(\frac{1}{2} - z + \frac{1}{\kappa}\right).$$

For the range of z for which case 2 holds this optimal price will give the maximal level of revenue r_2^*

$$\begin{aligned} r_2^* &= p_2^*\left(\frac{1}{2} - z + \frac{1 - p_2^*}{\kappa}\right) \\ &= \frac{1}{16\kappa}\left(-\kappa + 2d\kappa - 2\right)^2. \end{aligned}$$

Again, we need to determine the range of z such that charging p_2^* will give rise to the conditions for case 2. That is, for which values of z, p_2^* does indeed maximize revenue at the level given by r_2^*. The first condition for case 2 is that $w(z, p_2)$ does not lie entirely outside the market. That is

$$z - \frac{1 - p_2^*}{\kappa} = \frac{1}{4\kappa}\left(2d\kappa - 2 + \kappa\right) \leq \frac{1}{2}$$

rearranging gives

$$z \leq \frac{1}{2} + \frac{1}{\kappa}.$$

So case 2 applies up to case 0. We now determine the other end of the interval on which case 2 holds. The second endpoint is given by the second condition for case 2, namely that

$$\frac{1}{2} \leq z + \frac{1 - p_2^*}{\kappa} = \frac{1}{4\kappa}\left(6d\kappa + 2 - \kappa\right)$$

rearranging

$$\frac{1}{2} - \frac{1}{3\kappa} \leq z.$$

It only remains to find the optimal price and associated maximal level of revenue for the remaining range of z, $\frac{1}{2} - \frac{1}{2\kappa} \leq z \leq \frac{1}{2} + \frac{1}{3\kappa}$. This is the range of z that lies between cases 1 and 2. This case, which we will call case 3, is shown in Figure 4.6.

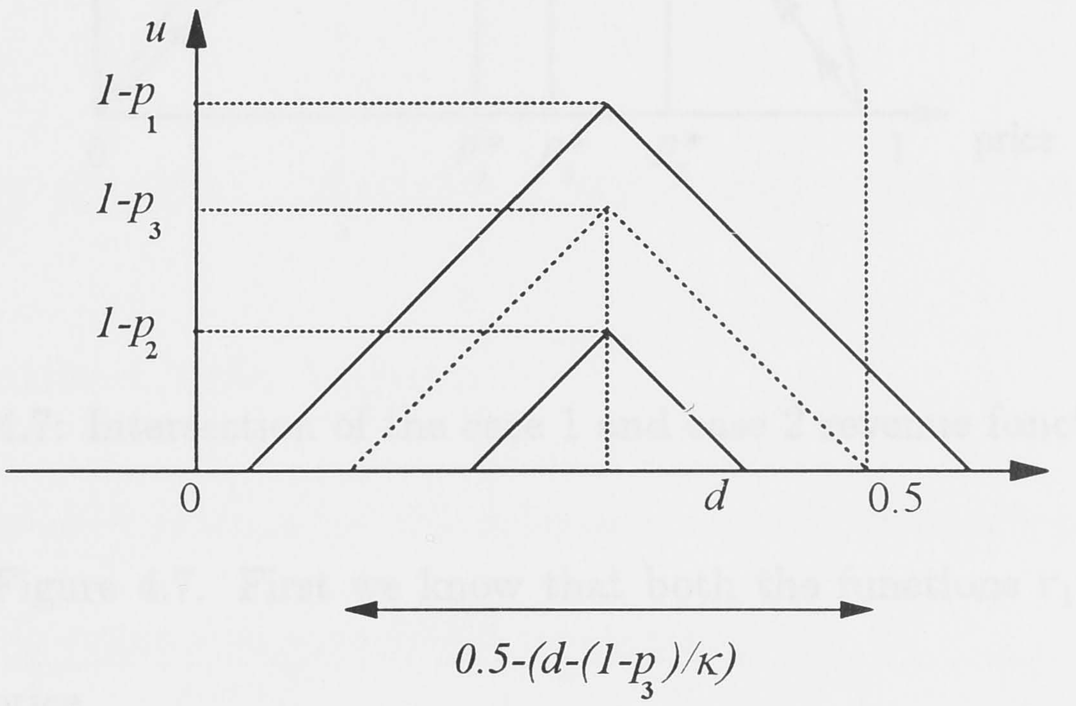


Figure 4.6: Case 3

If p_1^* is charged for z in this interval then by definition z is too close to the edge of the market for $w(z, p_1^*, \kappa)$to lie entirely within the market. Thus if p_1^* is charged, the conditions of case 1 are violated so it will not be the optimal price. Similarly charging p_2^* leads by definition to $w(z, p_2^*, \kappa)$ lying entirely within the market, violating the conditions of case 2 for which p_2^* is the optimal price.

We will now show that p_3 such that $z + \frac{1-p_3}{\kappa} = \frac{1}{2}$ is the revenue maximizing price for case 3. This is the price such that the right hand end of $w(z, p, \kappa)$ just touches the edge of the market.

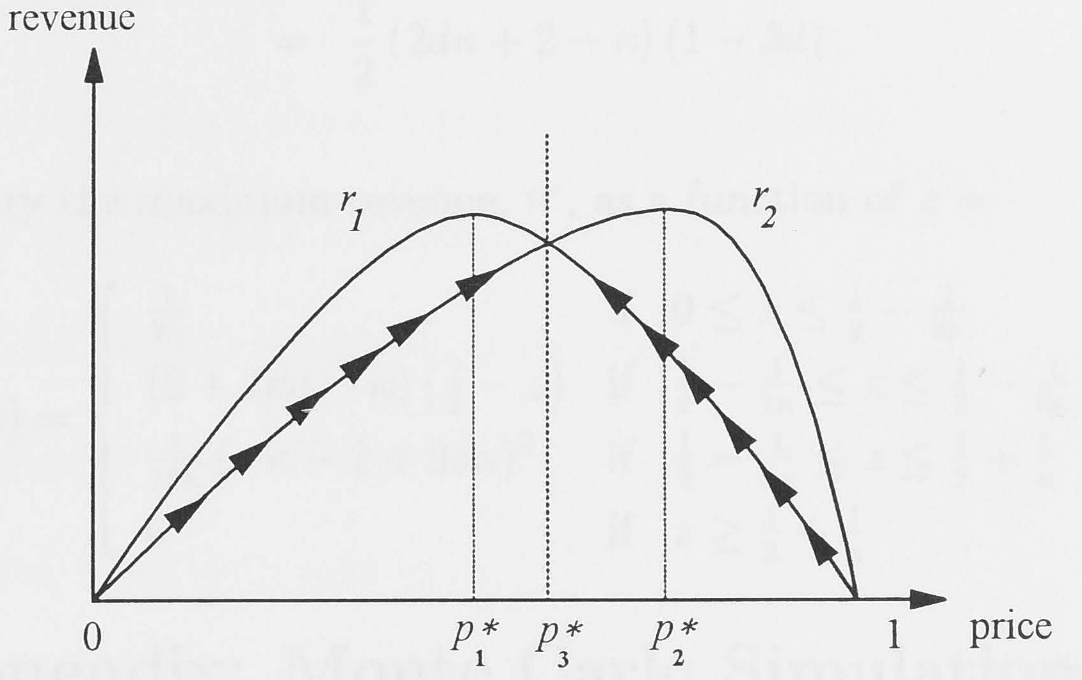


Figure 4.7: Intersection of the case 1 and case 2 revenue functions.

Consider Figure 4.7. First we know that both the functions r_1 and r_2 are quadratic in price.

On the interval $[0, p_3]$ we have, by construction, that $z + \frac{1-p_3}{\kappa} \geq \frac{1}{2}$, therefore on this interval, revenue is correctly given by r_2. However as was shown above (and is illustrated in Figure 4.7) r_2 is quadratic and its maximum lies to the right of p_3. Therefore revenue must increase as price is increased from 0 to p_3. Similarly on the interval $[p_3, 1]$ revenue is correctly given by the function r_1. Which implies that revenue increases as the price decreases from 1 to p_3. Hence p_3^*, the optimal price under case 3, is given by

$$p_3^* = \left(z + \frac{1}{\kappa} - \frac{1}{2}\right)\kappa$$

and the associated maximal level of revenue is

$$\begin{aligned} r_3^* &= p_3^* \left(2\frac{1-p_3^*}{\kappa}\right) \\ &= \frac{1}{2}\left(2d\kappa + 2 - \kappa\right)\left(1-2d\right). \end{aligned}$$

In summary the maximum revenue, r^*, as a function of z is:

$$r^*(z) = \begin{cases} \frac{1}{2\kappa} & \text{if } 0 \le z \le \frac{1}{2} - \frac{1}{2\kappa} \\ (2 + 2\kappa d - \kappa)\left(\frac{1}{2} - z\right) & \text{if } \frac{1}{2} - \frac{1}{2\kappa} \le z \le \frac{1}{2} - \frac{1}{3\kappa} \\ \frac{1}{16\kappa}\left(-\kappa - 2 + 2d\kappa\right)^2 & \text{if } \frac{1}{2} - \frac{1}{3\kappa} \le z \le \frac{1}{2} + \frac{1}{\kappa} \\ 0 & \text{if } z \ge \frac{1}{2} + \frac{1}{\kappa} \end{cases}.$$

4.8 Appendix: Monte Carlo Simulations

4.8.1 Monte Carlo Values

Values for expected revenue for the delay-is-everything hierarchies, calculated by monte Carlo simulation with 10000 repetitions.

	$\sigma = 0.2$		$\sigma = 0.075$	
$\log_2 N$	$\kappa = 20$	$\kappa = 2$	$\kappa = 20$	$\kappa = 2$
1	0.02101189	0.20356559	0.02354468	0.21894821
2	0.02124086	0.20751456	0.02446284	0.23685974
3	0.02037817	0.19973162	0.02479172	0.24273685
4	0.01941837	0.19143193	0.02486382	0.24453731
5	0.01826924	0.18082125	0.02485755	0.24444690
6	0.01722225	0.17138016	0.02479782	0.24313296
7	0.01631026	0.16327122	0.02464005	0.24116235
8	0.01541933	0.15566197	0.02451229	0.23909901
9	0.01477746	0.14930366	0.02435879	0.23722819
10	0.01417314	0.14355251	0.02412833	0.23444457
11	0.01369993	0.13899398	0.02392792	0.23217054
12	0.01330436	0.13499591	0.02369349	0.22945997

4.8.2 Code for the Monte Carlo Simulations

The Monte Carlo simulations were performed in Matlab. An example of the code is shown below:

```
%This program runs a monte Carlo simulation to find
%expected market shares for specific k and sigma
%values. The variable price monopolist model is used.
%The number of delays is everything, one  shot efficient
%hierarchies, is set with maxdelay.
nruns=input('number of replications=');
maxdelay=input('the maximum delay of a hierarchy=');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%  k=2 standard_deviation=0.75
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
k=2;
standard_deviation=0.075;
for run=1:nruns
%  Generate the observations, shocks and forecasts
observations=rand(2^(maxdelay-1),1)-0.5;
shocks=standard_deviation*randn(maxdelay,1);
for index=1:maxdelay
 forecasts(index)=mean(observations(1:2^(index-1)));
```

```
    actual_means(index)=sum(shocks(1:index));
end %index for loop
errors(run,1:maxdelay)=forecasts-actual_means;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%This loop works out the profit for a particular
%generated forecast error
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for index2=1:maxdelay
    if abs(errors(run,index2)) <= 0.5-(1/(2*k))
        revenue(run,index2)=1/(2*k);
    elseif abs(errors(run,index2)) > 0.5-(1/(2*k))
& abs(errors(run,index2)) < 0.5(1/(3*k))
        revenue(run,index2)=(2+2*k*abs(errors(run,index2))-k)*
        (0.5*abs(errors(run,index2)));
    elseif abs(errors(run,index2))>=0.5-(1/(3*k))
& abs(errors(run,index2))<0.5+(1/k)
      revenue(run,index2)=(1/(16*k))
*((-k-2+2*k*abs(errors(run,index2)))^2);
    else
        revenue(run,index2)=0;
    end  %if statement
end   %index2 for loop
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
end    %run loop
mrevenue=mean(revenue);
save k2s075 mrevenue;
```

4.9 Appendix: Comparison of Monte Carlo and Moment Approximation Methods

Figures 4.8 and 4.9 show the expected revenue predicted by both Monte Carlo simulation and the Normal approximation for various parameter values. The model used is the fixed price model of the previous chapter with the delay-is-everything hierarchies from this chapter.

In both figures the expected revenue curves produced by each method are very close. In particular it can be seen that the main results from the proceeding analysis will hold under either method.

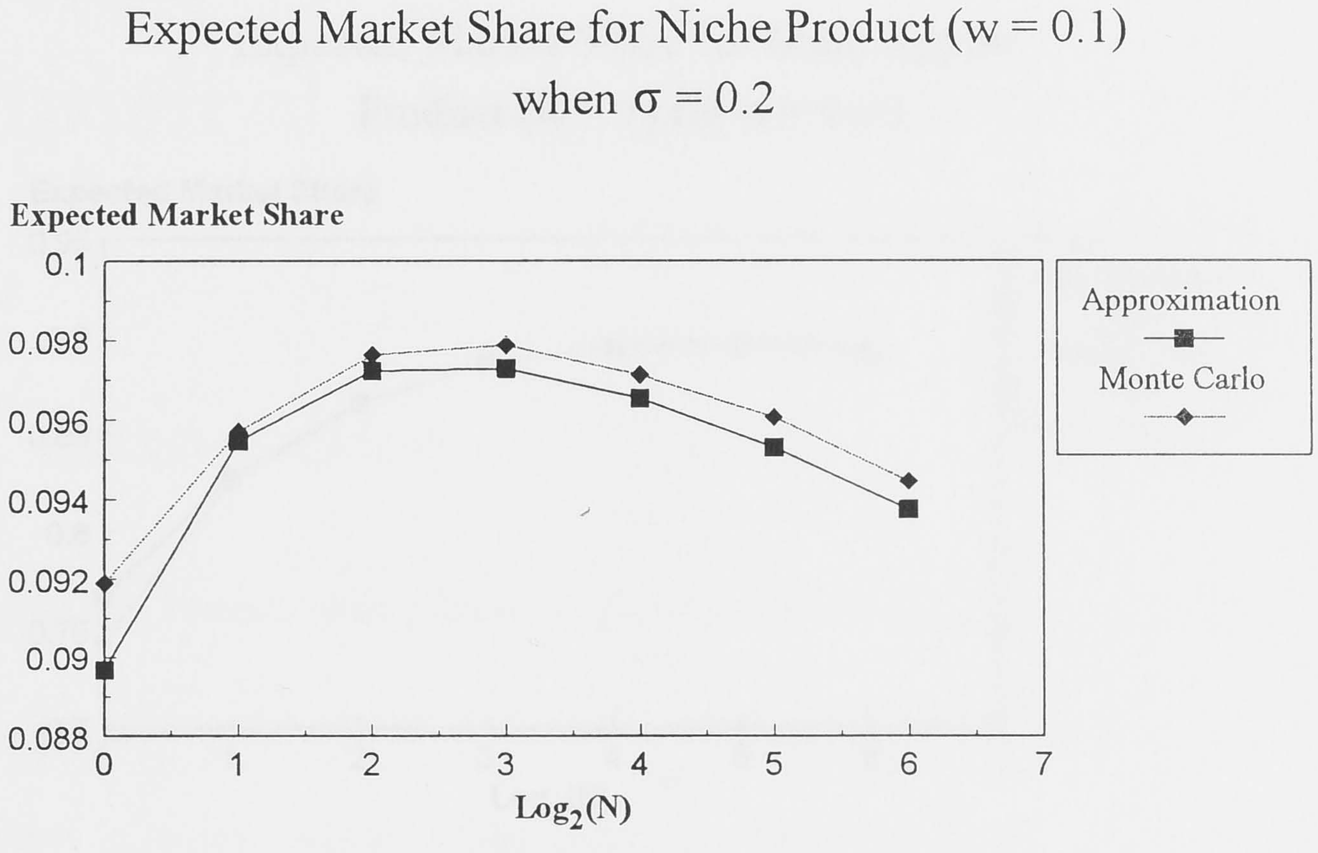


Figure 4.8: Comparison of Monte Carlo results and approximation with the Normal distribution.

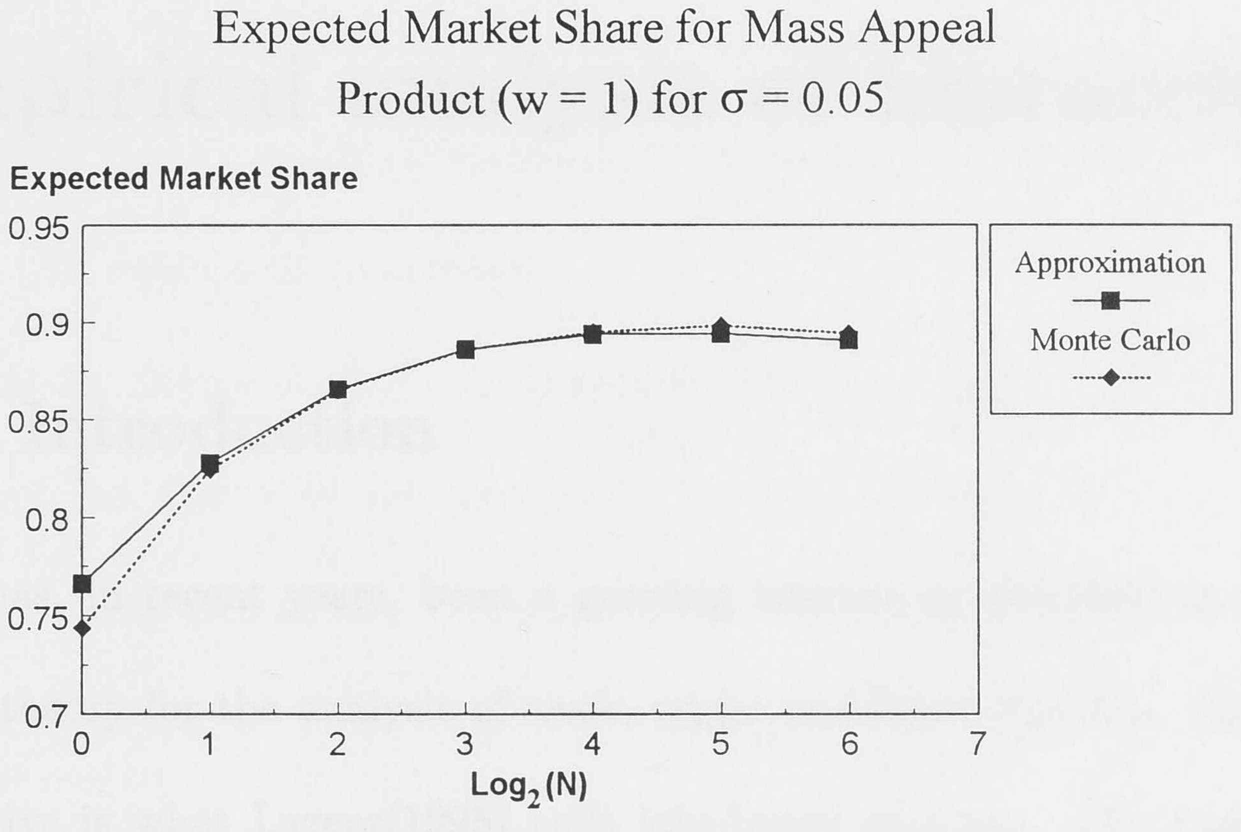


Figure 4.9: Comparison of the Monte Carlo results and the Normal approximation.

Chapter 5

Empirical analysis of hierarchies

5.1 Introduction

There has, in recent years, been a growing interest in alternatives to human capital theory for the analysis of work, wages and labor markets. The leading alternative is what Lazear(1995) calls jobs-based analysis. Jobs-based analysis is frequently identified with the theory of internal labor markets and the related empirical work on personnel records by Baker et al(1994a,1994b) and Lazear(1992). This work has used case studies of personnel records to identify a jobs hierarchy (career ladder) and to look at the patterns of promotions and wages associated with these hierarchies. Human capital theory is a supply motivated, closed form approach that interprets wages as the returns to investments by individuals. The jobs-based approach instead focuses on the connection between jobs and wages.

Lazear(1995) identifies a number of important theories in the jobs-based approach: tournaments, hierarchies, hedonic wage analysis, job investment, insurance and work-sharing. The least studied of the theories, in the context of labor markets, is the theory of hierarchies. The purpose of this paper is to present a brief survey of some hierarchy models and show their empirical predictions. The hierarchies approach then provides a framework in which the empirical analysis is conducted. The results of the empirical analysis can help to select between the existing models and more generally to indicate the facts which the ongoing work in this field needs to address.

Hierarchy theory analyses management structures and is primarily an outgrowth of the theory of the firm. The hierarchies considered are defined by decision making and control in firms and are therefore different, and more readily observed, than the jobs hierarchies considered by internal labor markets.

To see this difference, consider a professional sports team. The management hierarchy is quite clear: at the bottom are the players and above them is a coach. Above the coach there is probably a manager who would have other subordinates, besides the coach, dealing with such things as the stadium, merchandising and legal matters.

The jobs hierarchy is not so obvious because it is defined by patterns of promotion. Baker et al(1995,p256) identify two ways to define promotions:

"Lazear(1992) resolves this dilemma by defining promotions as move-

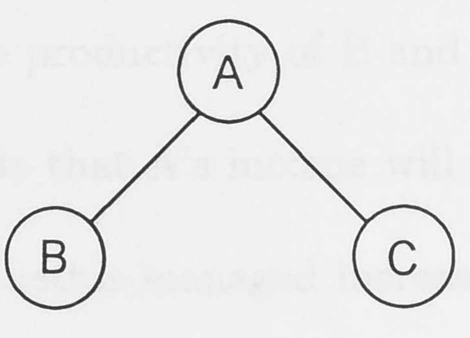
> ments from a job title with a lower average pay to a job title with a higher average pay....In our own work we used yearly patterns of job transitions to infer promotions".

Returning to the example of a sports team, the salaries of coaches are generally lower than those of top players so the Lazear approach would actually find that players who become coaches are demoted. This is less of a problem under the approach of Baker et al, but players only infrequently become coaches so Baker et al are more likely to find no hierarchical relationship between players and coaches.

This example is not intended as a criticism of the construction of job ladders but rather to highlight an important difference between job ladders and management hierarchies. To construct a firms job ladder requires analysis of the whole firm over a period of time, since the job ladder is defined by the dynamic process of promotion. Self-reporting of promotions is unlikely to be of use because the appropriate criteria are not simply change in job title or wage. On the other hand the management hierarchy for a firm can be defined at a point in time by mapping out the superior-subordinate relationship. The simplification goes further because the position of an individual in a hierarchy can be calculated just by asking her about the number of levels of subordinates below her and the number of levels of superiors above her. These are questions that can easily be asked in a traditional random sample of individuals.

Why job ladders are of interest is well known, but the significance of management hierarchies is less so. The functional relationship between a superior and a subordinate in a management hierarchy is a relationship affecting productive activity that has implications for wages not suggested by the job ladders approach. The actions of a supervisor spill over onto all her subordinates affecting their productivity and hence the wages of the subordinates in a manner which is determined by the type of spillover and the nature of the employment relationship. To see this more clearly consider the three simple organizations in Figures 5.1 and 5.2.

(a) Market



(b) 2 level hierarchy

Figure 5.1: Two possible organizations of three people.

Economic activity can either be organized across a market, Figure 5.1(a) or within a firm, Figures 5.1(b) and 5.2. We know from the transaction cost literature that the firm will be the chosen form of organization when it is more efficient. For a firm to be more efficient means, for example, that for the same effort the value of total output of the three individuals A,B and C is greater when they coordinate their activities by organizing themselves in a firm/hierarchy (5.1(b)) than when they act individually (5.1(a)). If the value of total output is higher then there is a surplus, generated by organizing in the hierarchy, which is to be split between those in the hierarchy.

The way in which the surplus is split will depend on the bargaining power and the threat points (outside options) of the various individuals involved. It is widely accepted that A, the boss of the two level firm will capture a large part of this surplus. But the situation is more complex because we have not yet considered the productive affect of A on B and C; and the possibility of other organizational forms.

In the two level hierarchy, A is the owner of the firm and manages B and C. If the actions of A affect the productivity of B and C, and hence total output, then it is natural to conclude that A's income will be increasing in A's ability to manage. However being better managed increases the marginal product of B and C, therefore if they have any bargaining power their pay will also be increasing in the quality of management provided by A. Other factors which will influence the outcome of the bargaining are the alternative options. Can B and C work in another firm or is self employment (market organization) their only option? Must A be the owner manager or could A be a middle manager in a larger firm?

Now consider the three level hierarchy in Figure 5.2. The actions of D influence the productivity of the four bottom level workers and the two managers, so it might be reasonable to conjecture that D's income is even higher than A's was in the two level hierarchy. A is no longer the boss but is still the manager in charge of B and C. What influence would this change in hierarchies have on the incomes of A,B and C? The answer to this question depends on exactly how management effects productivity and wage determination.

These simple examples raise two questions. How do wages change with the number of levels of subordinates below an individual? An example of this would be comparing the wage of say A and B (B is one level below A) or B and D (B is two levels below D) in the three level hierarchy. How does the management structure in which an individual works affect their wage (for example comparing the wage of B in the two level hierarchy with B's wage in the three level hierarchy)?

These are empirical questions that could be answered with an individual level survey which includes the appropriate questions about the individuals place in the hierarchy. Fortunately such a survey exists, which will allow the empirical analysis of these questions for the first time.

The paper first provides a brief survey of some relevant results from hierarchy theory in Section 5.2. This provides a framework for the empirical analysis and a more formal motivation for considering the impact of hierarchies on wages. The models chosen show some of the range of predictions possible from hierarchy

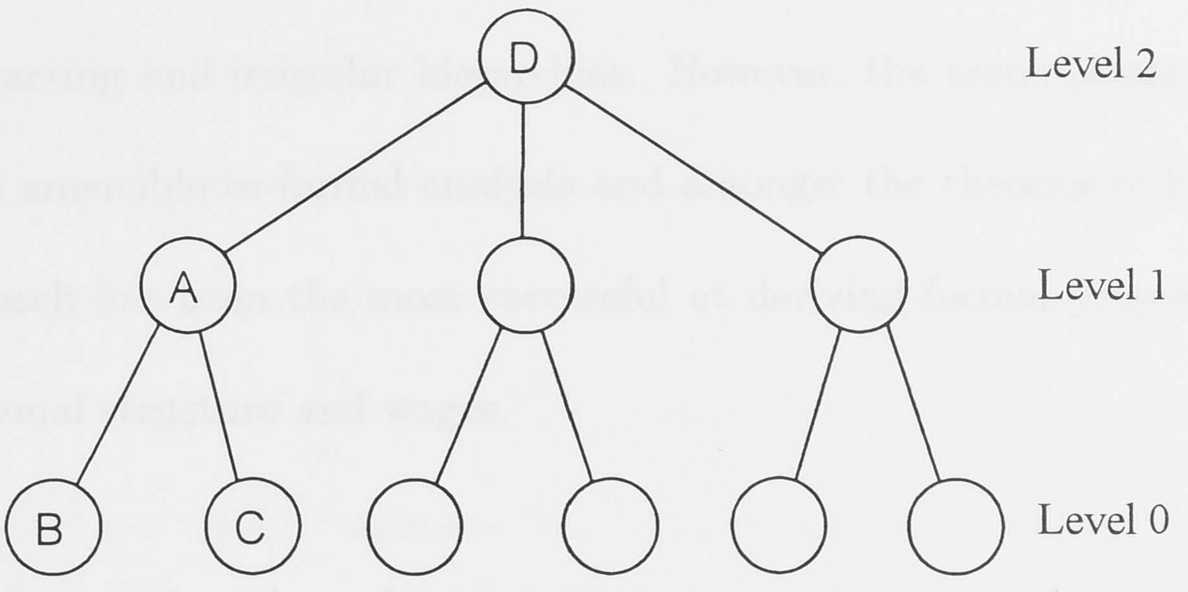


Figure 5.2: A three level hierarchy.

theory. The data set used is described in Section 5.3. The data set has both human capital variables and hierarchy variables so the empirical analysis in Section 5.4 utilizes both theories.

The paper first provides a brief survey of some relevant results from hierarchy theory in Section 5.2. The data set used is described in section 5.3. The data set has both human capital variables and hierarchy variables so the empirical analysis in Section 5.4 utilizes both theories.

5.2 The Grand Contract Model

The grand contract model is so named because the owner/manager of the firm chooses the incentive structure for the whole organization and no other contracts are written.

The assumptions of the grand contract model are highly stylized and ignore a number of significant factors such as endogenous choice of hierarchy, side contracting and irregular hierarchies. However, the assumptions do make the model amenable to formal analysis and amongst the theories of hierarchies this approach has been the most successful at deriving formal propositions on organizational structure and wages.

5.2.1 Organizational structure

Consider a firm that has a hierarchical organization. Its internal structure is akin to the common intuitive model of management organization expressed in organizational charts. Those at the bottom are occupied solely with the direct production of the firm's output. They have supervisors to coordinate and direct their work. The supervisors in turn are supervised, and so on, up the hierarchy to the owner or chief executive officer. For convenience the person at the top of the hierarchy will always be referred to as the owner. We assume that everyone except the owner has exactly one superior. Thus the organizational structure is a tree, with the owner at the root and the workers at the leaves.

Each worker is assumed to have exactly the same number of superiors in the chain between the worker and the owner, and anyone who does not have subordinates is a worker. Hence we can uniquely define the level of each person in the tree. All workers are at level 0. Those who directly supervise workers are at level 1, those who supervise level 1 supervisors are at level 2, and so on up the hierarchy until the boss who is at level L.

The number of employees in level l is denoted x_l and must be a positive integer for $l \leq L$. By definition $x_L = 1$. The span of control for a supervisor is the number of people in the level below her who she directly supervises. It is assumed that the span of control is constant across supervisors at the same level. Hence the span of control at level l, denoted s_l, satisfies $x_{l-1} = s_l x_l$.

The owner is a residual claimant and hence maximizes profits by maximizing revenue less the wage payments to employees. The determinants of revenue and wages are considered next.

5.2.2 Production and Management

Production is undertaken by the N workers in level 0. N is fixed for the derivation of the optimal organizational structure. Hence variations in organizational structure all occur in the supervisory levels. Qian(1994) explains the assumption of a fixed N by assuming a technology which requires a fixed amount of capital, k, per worker and that the firm has a fixed capital stock K. Qian does not add a capital market in which K is determined, so this additional assumption adds nothing substantial to the model and will be omitted in the following analysis.

Only a single period of time is considered so the model is static.[1]

Two possible classes of activity are envisaged for supervisors in the grand contract models: monitoring and management. Monitoring is checking the work of direct subordinates in order to prevent shirking. Monitoring is considered, along with efficiency wages, in the following section. The management function is more subtle. It is an attempt to capture the planning and coordination tasks which are a significant part of what supervisors do.

There is a division in the literature between those that allow for some type of management function, for example Williamson(1967), Beckmann(1977), Rosen(1982) and Qian(1994) and those that do not, for example Calvo and Wellisz(1979). In those papers that do include a management function this is achieved by defining a recursive production technology. We will attempt to define an encompassing framework that allows both the comparison of the various specifications of a management function and a comparison with those models which do not include a management function.

Let α_l be the effective skill that employees at level l apply to their jobs. Effective skill is in the range $0 \leq \alpha_l \leq 1$ where 1 is the maximum effective skill possible. α_l is constant within a level but may vary between levels. Effective

[1]Clearly in the real world the actual number of people employed must be an integer. For the purpose of analysis this restraint is often ignored, see for example Qian(1994). In a richer model a fraction of an employee would be possible by considering a partime employee as a fraction of a fulltime employee. Meagher(1996b) and Van Zandt(1996) consider these different types of employment in the information processing model of teams.

skill is a composite open to various interpretations. It could be an individual's ability (either acquired or inherent) or effort, or some combination of both. Williamson(1967) assumes that $\alpha_l = a$, where a is constant and $0 < a < 1$.

Qian fixes $\alpha_L = 1$ and makes α_l a choice variable for all other employees.

Let $g(\alpha)$ be the opportunity cost of supplying α (for those individuals for whom α is possible).

Following Qian(1994) the recursive technology of management is defined by assuming that an intermediate product called "managerial effectiveness" is produced. Planning and coordination which are the intuitive command functions on which managerial effectiveness are based flow down a hierarchy. Hence a supervisor at level l produces managerial effectiveness by taking y_{l+1}, the managerial effectiveness of their immediate supervisor, and combining it with their effective skill α_l to produce their managerial effectiveness y_l. This process can be expressed by the function $y_l = F_l(y_{l+1}, \alpha_l)$. For simplicity F_l is generally assumed to be invariant across levels and to have a simple functional form.[2]

Williamson(1967) and Qian(1994) both assume a multiplicative form $y_l = y_{l+1}\alpha_l$. It follows that in Williamson(1967) $y_l = a^{L+1-l}$ and Qian(1994) $y_l = \alpha_l \alpha_{l+1} \dots \alpha_L$.

The maximum output for a worker is $\theta > 0$, which occurs when $y_1 = 1$. Therefore total/gross output from all workers is $\theta N y_1$, which is also total revenue since price is normalized to 1. This simple model makes output a func-

[2]It is not known what properties $F_t(.)$ must have in order for the model to be tractable.

tion of the production technology, the effectiveness of production workers, and the effectiveness of management.

5.2.3 Monitoring

Supervisors in the hierarchy also perform monitoring on the inputs and outputs of direct subordinates in order to detect shirkers.

To encompass the variety of modelling approaches we define β_l to be effective skill in supervision for a level l employee. Similar to α_l, effective skill in supervision is a composite of ability and effort. Again it is assumed that β_l is constant within a level but may vary between levels.

Monitoring is assumed to be the only activity that supervisors undertake which requires time. Shirking is detected by observing the work of a direct subordinate, and hence only one subordinate can be monitored at a time. The effectiveness of a supervisor in observing the effort of a subordinate is determined by the supervisor's β (effective skill in supervision). Thus the probability P_l of a level l supervisor detecting a shirking subordinate at level $l-1$ is β_l/s_l.

Let $h(\beta)$ be the opportunity cost to an employee to supplying β (assuming that they are able to).

5.2.4 Efficiency Wages

In all the following cases limited liability is assumed so that wages cannot be negative. In order to implement an effective managerial skill of α^* the following family of incentive schemes is considered:

pay w if $\alpha \geq \alpha^*$ is known, or if α is not known; and

pay 0 if $\alpha < \alpha^*$ is known.

The predictions for wages in hierarchies from the grand contract model depend on the particular specification of the model. In the following, a range of results on wages are surveyed.

Williamson(1967)

Williamson(1967) pioneered the study of management hierarchies with his model of loss of control in supervision. However wages are assumed to take the form $w_l = w_0\eta^{L-l}$ where $\eta > 1$. Thus the only empirical test concerning wages that can be made of the Williamson model is this assumption. The model has no implications for changing wages with firm size, or indeed for organizational structure to determine the level of wages within a level.

Calvo and Wellisz(1979)

Calvo and Wellisz(1979) focus on income distributions and ability. The following assumptions are particular to their model. The probability, P_l, of a level l worker getting caught shirking is

$$p_l = \frac{\beta_{l+1}}{s_{l+1}}.$$

That is, the probability of getting caught depends not just on the number of other employees being supervised but also on the proficiency in supervision. It is assumed that the owner's proficiency is 1.

This model focuses on supervision alone, that is it assumes $\alpha_l = 1$ and hence $y_l = 1$ for all individuals who are working and $\alpha_l = 0$ for all individuals who shirk. Thus employees choose either full effort in work or complete idleness. Let the welfare difference between effort and idleness be $k = g(0) - g(1)$. Then the efficiency wage under the given set of contracts must satisfy

$$w_l \geq p_l h(\beta_l) + (1 - p_l)(w_l + k)$$

hence

$$w_l = (1/p_l - 1) k + h(\beta_l).$$

Calvo and Wellisz explicitly assume that the outside option for an individual is self-employment, which they implicitly take to involve full effort: $\alpha = 1$. Thus the opportunity cost term $h(\beta_l)$ is defined net of $g(1)$ in their analysis.

The two relevant results from Calvo and Wellisz(1979) are not stated explicitly in propositions in their presentation. Thus I quote how they describe their results. Their first result is the most contentious so I have reproduced their derivation of the result in Appendix A of this chapter. Their second result is much more standard so I refer the interested reader to the original paper for the derivation.

The first important empirical implication of the Calvo and Wellisz(1979) model is the following:

Conjecture 10 *The optimal wage and supervisory effectiveness for a level is independent of the number of levels above that level.*

> "...the optimal wage and labor quality for a hierarchic layer is independent of the number of layers superior to the one for which the choice is made. ...we have just shown that the production workers will be offered the same conditions of work regardless of the number of hierarchic layers in the enterprise, and the same will hold for the second- and higher -layer employees. The result is realistic, and, unlike some results that follow, it is reassuringly neoclassical."

The empirical validity of this result is not nearly as self evident as Calvo and Wellisz assert, as will be shown in the empirical testing in Section 5.4. Their second major finding corresponds much more closely with casual empiricism:

> "...even if work at all hierarchic levels is equally difficult and equally onerous, a profit-maximizing enterprise will assign the better quality workers to the higher ranks and will pay them higher wages than those paid at the lower ranks to the lower-quality workers."

They also prove that the differences in ability of workers at different levels is not the only factor causing inter level wage differentials:

"...which shows that the interlayer wage differentials are greater than the differentials in effective labor per physical worker. In a hierarchic organization there is, as it were, a multiplicative productivity effect. If a worker shirks, the firm loses the worker's product. If a supervisor shirks and, as a consequence, the workers under him shirk, too, the firm loses the produce of the entire productive workers' team. This is the basic reason for assigning the more productive workers to the higher-level jobs and offering them a wage higher than would be accounted for by their higher efficiency."

This is an interesting theoretical result and can be expressed as the following empirical conjecture:

Conjecture 11 *Wages should increase with hierarchical level after actual ability is corrected for, not just observed characteristics.*

Qian(1994) is critical of the approach in Calvo and Wellisz on two fronts. First the choice of effort is binary: work or shirk. This assumption is relaxed in Qian's framework, which is considered in the next section. Qian also argues that Calvo and Wellisz assume a fixed number of levels in the hierarchy and that this causes a bottle neck, at the top of the organization, which produces the above results. Qian's bottleneck criticism appears unfounded since the first result from Calvo and Wellisz showed that the number of levels above a level is irrelevant.

Qian(1994)

Qian(1994) assumes that $\beta = 1$ and $h(\beta) = 0$ for all employees. Incentive compatibility requires that

$$w - g(\alpha^*) \geq P \cdot 0 + (1 - P) w - g(\alpha), \text{ for all } \alpha < \alpha^*.$$

This gives an efficiency wage of $w = g(\alpha^*) / P$. Now $P = 1/s_l$ if the employee under consideration is at level $l - 1$, thus the wage function for each level of the hierarchy is given by

$$w_l = g(\alpha_l) s_l.$$

The optimization problem of the organization is expressed by

$$\max_{s_l, a_l, L} \theta N y_1 - \sum_{l=1}^{L} \left(g(\alpha_l) s_l x_l\right)$$

$$\begin{aligned} s.t. \quad x_l &= s_l x_{l-1} \\ y_l &= y_{l-1} a_l \\ x_0 &= 1, x_L = N \text{ and } y_0 = 1. \end{aligned}$$

Qian(1994) solves this problem for the case where L and s_l are integers, and for a continuous approximation where all the variable can take on continuous values.

The continuous approximation uses a dynamic programming approach based on Keren and Levhari (1979). Under the continuous approximation the firm's optimization problem becomes

$$\max_{s_l, a_l, L} \theta N y_1 - \int_0^L \left(g(\alpha_l) s_l x_l\right) dl$$

$$s.t.\ \dot{x}_l = x_l \log(s_l)$$

$$\dot{y}_l = y_l \log(\alpha_l)$$

$$x_0 = 1, x_L = N \text{ and } y_0 = 1.$$

For the purpose of the empirical analysis in Section 5.4 the following results from Qian(1994, p537-538) are relevant.

Proposition 12 *(i) As N increases, the total number of tiers of the hierarchy increases ($dL/dN > 0$); and*

(ii)as the productivity parameter θ increases, the total number of tiers of the hierarchy increases if and only if the span of control at the bottom is larger than that at the top.

Proposition 13 *For any person remaining in level l, $0 \leq l < L$, when N increases, the optimal hierarchy will adjust so that his wage payment decreases.*

It follows from Proposition 12 that an individual remaining at the same level, as N increases, will have more levels above them in the hierarchy. From Proposition 13 it follows that the wages of such a person will decrease. This suggests the following testable conjecture.

Conjecture 14 *If hierarchies are optimal and they vary in height because of N, then for individuals at the same level we would expect increasing the number of tiers above the individual to have a negative effect on wages.*

Proposition 15 *For any person maintaining τ levels above them, $0 \leq \tau < L$, when N increases, the optimal hierarchy will adjust so that his wage payment increases.*

This leads to another testable conjecture:

Conjecture 16 *If hierarchies are optimal and they vary in height because of N, then for individuals with the same number of levels above them we would expect increasing the number of levels below the individual to have a positive effect on wages.*

Qian motivates treating N as fixed due to fixed capital stock K. This is more likely to be binding in the short run than in the long run. The conventional economic wisdom is that the optimal scale of production is chosen in the long run. This intuition suggests that differences in the size of hierarchies will occur as the result of differences in the values of θ. Note that Qian's analysis does not address the empirically relevant question of how the total number of employees varies as N and θ change. This is because the spans of control may also readjust as the height changes.

We turn now to the case where changes in the hierarchy and wages are due to changes in θ.

Proposition 17 *For any person with a fixed number of levels below or above them when θ changes, the optimal hierarchy adjusts so that their wage increases.*

By similar reasoning to that used above, Proposition 17 leads directly to the two following empirical conjectures.

Conjecture 18 *If hierarchies are optimal and they vary in height because of θ, then holding the number of levels below an individual constant while increasing the number of levels above them should be associated with increasing wages.*

Conjecture 19 *If hierarchies are optimal and they vary in height because of θ, then holding the number of levels above the individual constant while increasing the number of levels below should be associated with increasing wages.*

5.3 The Data

The data used in the following analysis is from four surveys of Australian citizens conducted by the National Social Science Survey (NSSS). The surveys are: NSSS First Round 1984, NSSS 1986-87: Role of Government, NSSS 1987-1988: Inequality and NSSS 1995-96: International Social Survey Programme-National Identity Module & Reshaping Australian Institutions. Sampling was chosen by stratified systematic random sampling of the Australian electoral roll.[3] The 1984 survey was conducted face to face while the other surveys used a self completion (mail out, mail back) method. Bean (1991) reports that the surveys closely approximate the Australian population as recorded in the 1986 census

[3] All Australian citizens are required to be on the electoral roll.

and that there do not seem to be any significant differences in the rate or quality of response between the different survey methods.

These four surveys are pooled and a sub-sample is used. The sub-sample consists of men between 18 and 65 years, who work at least 1250 hours a year, work in the private sector and are not self employed. Only those with earning of less than \$250000 and an effective hourly wage of at least \$3 are used.[4]

The definitions for the variables used are given in Table 5.1, with the exception of the above and below variables which require more careful exposition.

Table 5.1

Variable	Definition
earnings	Total pre-tax wage and salary income from the previous 12 months.
hoursworked	Usual number of hours worked per week for pay.
weeksworked	Number of weeks of paid work in the last 12 months (including paid vacation and sick leave).
lnwages	Natural logarithm of *earnings/(hoursworked*$\times$ *weeksworked)*.
education	Number of years of education, including imputed values for post secondary qualifications.
experience	Number of years worked for pay (where the individual worked most of the year).
experience^2	*experience*$\times$ *experience*.
married	Married or in a defacto relationship.
union	Trade union member.
metro	Lives in a city with a population of at least 500000.
firm size	One plus the number of individuals, beside the respondent, who usually work for the company as full time employees.
yearXX	Dummy variable for survey year.

The supervisory hierarchy in which the respondent works is calculated from

[4]The minimum hourly wage in Australia through out this period exceeded \$3. These two exclusions remove a small number of individuals, the majority of whom appear to have misreported.

the following four questions. Do you have a supervisor in your job to whom you are directly responsible? If so, does that person have a supervisor on the job to whom he is directly responsible? In your job do you supervise anyone who is directly responsible to you? Do any of these persons supervise anyone else? From these questions it is possible to calculate *below* and *above.*

below is the number of levels of subordinates below an individual. It takes on the values $0, 1$ and 2. Similarly, *above* is the number of levels of supervisors a respondent has, referred to as the individuals superiors. Again, it takes on the distinct values $0, 1$ and 2. It should be noted that for these two variables the value 2 does not mean exactly two levels (of superiors or subordinates, as appropriate) but rather at least two levels. *below* and *above* are then broken down into dummy variables for each of their respective values. For example, *below=1* equals 1 only for those respondents who have exactly 1 level of subordinates, that is when below equals 1. To see how these variables represent hierarchical position, consider Figures 5.1 and 5.2. A is at the bottom of the hierarchies in both figures so below is 0 for A in both cases, hence *below=1* is 0 and *below=2* is 0. In Figure 5.1 there is one level above A so *above* takes the value 1 (*above=1* is 1 and *above=2* is 0). However in Figure 5.2 there are two levels above A so *above* takes the value 2 for A in this case (*above=1* is 0 and *above=2* is 1).

The means and standard deviations for the variables are shown in Table 5.2.

Table 5.2

Means and Standard Deviations of variables for the sample of men in full-time employment used in the empirical analysis

Variable	**Mean**	**Std. Deviation**
lnwages	2.463	0.520
year86	0.185	0.388
year87	0.215	0.411
year95	0.2615	0.440
education	11.555	4.700
experience	21.053	12.210
experience^2	592.163	585.101
married	0.693	0.461
metro	0.504	0.500
union	0.359	0.480
below=1	0.318	0.466
below=2	0.235	0.424
above=1	0.213	0.410
above=2	0.575	0.495
married×year95	0.194	0.396
union×year95	0.076	0.265
firm size: 10-99	0.316	0.465
firm size: 100+	0.501	0.500

5.4 Empirical Analysis

Table 5.3 shows the means of real annual earning by hierarchical position, for the sample. The pattern in mean wages for supervision matches with casual empiricism: average earnings increase with levels in the hierarchy below the individual. Earnings increase from \$24956 for workers at the bottom of the hierarchy, to \$30215 for supervisors, and \$39145 for those at higher levels.

The pattern in average yearly earnings, when they are tabulated by the number of levels above, is less clear. The pattern that we observe is that those with 2 or more levels of bosses get paid more that those with only 1 level of bosses (for below equals 0 or 1), but get paid less when there are 2 or more levels below.

Table 5.3 fails to control for a number of other factors, such as education and experience. To allow for these other factors a more detailed empirical analysis is conducted in the rest of this section.

Table 5.3

Means and frequencies of real annual wage and salary income by hierarchical position for men working full-time (base year is 1987).

Levels above respondent		**Levels below respondent** 0	1	2 or more	Total
0	Mean	$24898	$33626	$41600	$32612
	Frequency	76	73	57	206
1	Mean	$22911	$25189	$38565	$27260
	Frequency	81	81	46	208
2 or more	Mean	$25566	$31229	$38245	$29989
	Frequency	279	156	126	561
Total	Mean	$24956	$30215	$39145	$29961
	Frequency	436	310	229	975

5.4.1 Estimating Equation

We turn now to an empirical examination of the relationship between hierarchies and income. The theory surveyed in Section 5.2 did not suggest a particular functional form to estimate so the widely used human capital wage equation is utilized as the framework for the empirical analysis.

Using annual income as the dependent variable would capture the effects of both how much people work and the rate at which they are paid. It is an empirical fact that hours do vary, however in the theoretical models everyone works the same amount of time (one period). Thus it is more natural to use wages (hourly wage and salary income) rather than annual income.

Ability and training are also likely to play an important role in determining wages. In the Calvo and Wellisz model workers had different abilities, but there were hierarchical effects beyond the effects of ability. In the Qian model workers were of identical ability. Either way it is important to correct as much as possible for an individual's ability before examining the effects of hierarchies.

A leading approach for modeling the link between wages and ability is human capital theory. Human capital theory gives rise to the widely used wage equation of equation 5.1.

$$\ln W_i = a + bS_i + cEXP_i + dEXP_i^2 + eZ_i + u_i \tag{5.1}$$

Where i denotes individual i, $\ln W$ is the natural logarithm of hourly wages, S is years of schooling, EXP is length of time in the labor force, Z is a vector of other wage determining variables and u is a randomly distributed error term.

The derivation of equation 5.1 using human capital theory is covered in Appendix B of this chapter. There are other theoretical explanations for the wage equation, but for our purposes it is sufficient that it is the leading tool for the analysis of wages as a function of an individuals characteristics.

The grand contract models suggest that the variables *below=1*, *below=2*, *above=1* and *above=2* should be included in the Z vector. The signs and significances of the various coefficients will determine which particular specification of the grand contract model matches with the empirical findings.

Theory developed elsewhere and other empirical studies suggest that the variables on firm size and the variables *married*, *metro* and *union* should also be included in the analysis. For a survey of empirical results from human capital theory applied to Australia see Preston(1997).

5.4.2 Regression Results

Basic Model

The results from the estimation of four models are shown in Table 5.5, the results for a selection of variables are summarized in Table 5.4. Log of nominal hourly earnings is used as the dependent variable in each regression. Survey results from four years have been used. In order to correct for inflation, a dummy variable is included for the year in which an observation was recorded. The omitted year is 1984.

Three of the surveys are from consecutive years in the mid 1980's, so there is no a priori reason to expect that the independent variables may change in their effect on log wages over such a short period. The fourth survey is from 1995, making it plausible that there may have been some change in the relationship between the independent variables and *lnwages*. To allow for this possibility, each of the independent variables (excluding the year dummies) is interacted with the dummy variable for 1995. This procedure was followed for a number of different model specifications and only two interactions, those for *married* and *union*, showed to be statistically significant in any of these models. Hence, *married*$\times$*year95* and *union*$\times$*year95* are included in each of the reported regressions. The results on hierarchies are robust to the inclusion and exclusion of *married*$\times$*year95* and *union*$\times$*year95*.

Model R5.1 is a basic wage equation plus dummy variables for being married, living in a large metropolitan area, and being a union member. These individual characteristics have frequently been shown in the work of other authors to be significant determinants of wages in Australia, see Preston(1997).

The coefficients in model R5.1 are similar to those commonly found in the literature on wages in Australia. For example, Preston (1997, p60) reports 0.043 and -0.000765 as the coefficient estimates for *experience* and *experience*^{2} (for males in the private sector), neither of which are statistically significantly different to the corresponding estimates in R5.1. The effects of hierarchy and firm size variables are considered next. Some interesting results not directly related to hierarchies are discussed briefly in section 5.4.2.

Table 5.4

Percentage change in wages as a result of one unit change in independent variable.

	Model			
	R5.1	R5.2	R5.3	R5.4
experience, evaluated at:				
experience = 5	3.06	2.65	2.89	2.59
experience = 10	2.47	2.13	2.33	2.08
experience = 20	1.29	1.09	1.21	1.06
Other variables				
education	6.84	6.12	6.22	5.70
below=1		9.25		8.54
below=2		21.9		18.0
above=1		-1.70		-1.74
above=2		10.1		2.61
firm size:10-99			15.1	12.1
firm size:100+			27.2	22.3

Hierarchy and Firm Results

Model R5.2 is Model R5.1 plus the variables on position in the hierarchy: *below=1*, *below=2*, *above=1*, and *above=2*. Holding everything else constant, wages are approximately 9% higher for those in level 1 than for those in level 0. Similarly, the wages of those in level 2 and higher are approximately 22% higher than those for level 0 workers (everything else constant). This shows that wages increase with position in the hierarchy, even after correcting for an individual's observed education and experience. This result is consistent with Conjectures 10, 16 and 18.

The above variables show that having two or more levels of superiors increases wages by 10% over those with either zero or one level of superiors. There is no significant difference between having one and zero levels of supervisors, everything else constant. These results for *above* and *below* indicate that both level in the hierarchy and height of the hierarchy have significant positive effects on wages.

Comparing Model R5.1 and Model R5.2, we see that the *education* and *experience* coefficients are smaller when we correct for hierarchy effects. However, these human capital effects are still positive and significant.

Overall, these two sets of regression results show that both human capital and hierarchical factors are important in determining wages.

Model R5.3 is Model R5.1 expanded to include firm size variables. Similar to other empirical findings, the firm size effect is positive and strongest for the largest companies. It is difficult to make direct comparisons with other studies because of different variable definitions and categories, however Miller and Mulvey (1996) also find approximately a 20 percentage point difference between the largest and smallest firms. Allowing for firm size also reduces the influence of the human capital variables. However, unlike the hierarchy variables, firm size reduces the coefficient on union significantly.

One problem with the analysis so far is that hierarchy and firm size effects are likely to be confounded because larger companies, in general, are more likely to have larger hierarchies. With the variables available, it is not possible to separate out the common influences. In Model R5.4 both the firm size and hierarchy variables are added to the simple Model R5.1 specification. This specification shows which effects are robust to the confounding problem.

Model R5.4 shows a similar pattern of returns to being a supervisor to that seen in Model R5.2. The effects in Model R5.4 are slightly smaller for *below=1* (0.7 percentage points approximately). For *below=2*, the decrease in the coefficient is much larger between Model R5.2 and Model R5.4. The coefficient decreases by 4 percentage points, which is an 18% decrease.

Neither *above=1* nor *above=2* is significant in Model R5.4. The firm size effects also decrease in both size and significance when the hierarchy variables are also included. The magnitude of the decrease is quite large, indicating that ignoring level in the hierarchy (as in Model R5.3) leads to an overestimation of the firm size effect by about 25%.

The overall effect of both the firm size and hierarchy variables on the human capital variables is quite large. Comparing the coefficient on *education* in Model R5.1 with Model R5.4 shows that ignoring organizational effects leads to a 20% overestimation of the influence of education.

Comparing the marginal effect of *experience* at the 5, 10 and 20 year levels, shown in Table 5.4, shows that the simple model R5.1 overestimates by approximately 20% on average, compared to R.54. Plots of the returns to experience, using the coefficients from Table 5.5, are shown in Figure 5.3.

In summary, there is clear evidence that both traditional human capital variables and height in the hierarchy are important determinants of wages. This is consistent with the hierarchy models of both Calvo and Wellisz, and Qian. The Calvo and Wellisz approach is more attractive since it models both skill and hierarchical position.

It is also clear that there is another effect due to the organization in which an individual works. The exact nature of the effect, or more likely effects, is not revealed by the analysis here. However, this effect is positively related to both firm size and the height of the hierarchy above an individual. The presence of this positive relationship, whatever its exact form, clearly rejects Conjecture 14 of a negative relationship, and also provides strong evidence against Conjecture 11 of irrelevance of the rest of the organization. This leaves Conjecture 18, of a positive relationship between the number of levels above an individual and wages, as the best supported by the data.

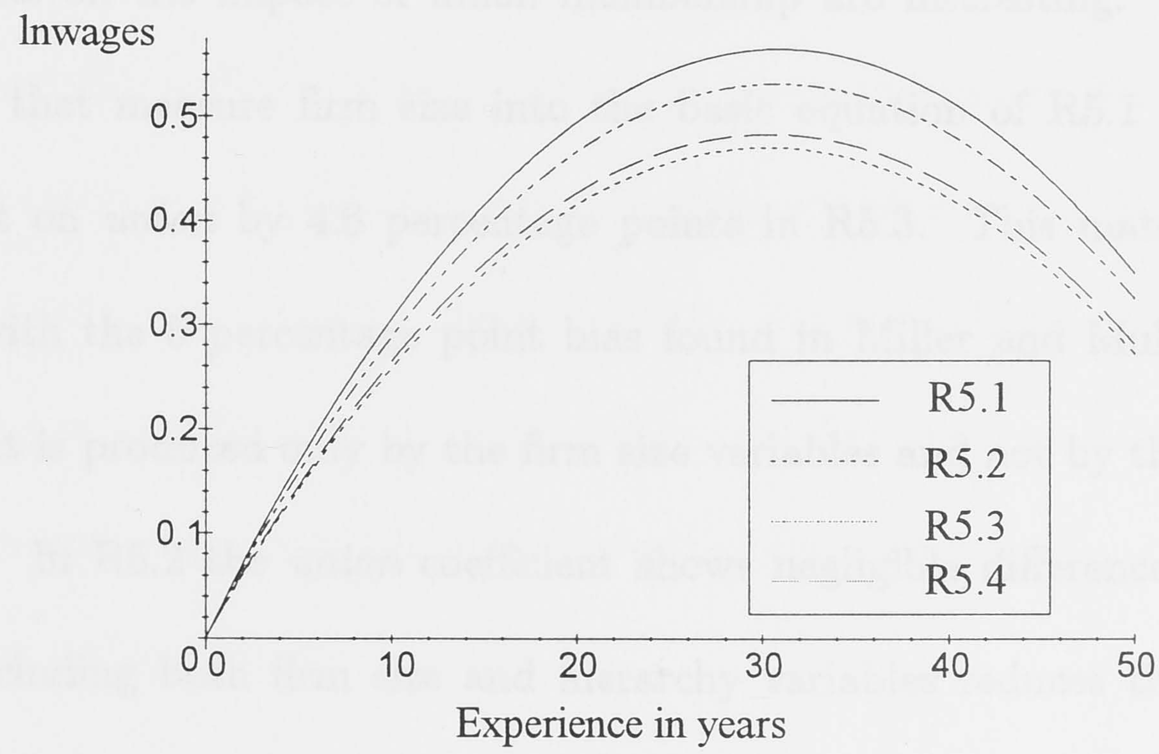


Figure 5.3: Effect on lnwages of different levels of experience for the four regression models.

Diagnostics

According to the Reset test, only the simple regression of R5.1 is mis-specified. Including either the firm size or hierarchy variables eliminates this mis-specification, again indicating the importance of organizational structure in determining wages.

All the models exhibit heteroscedasticity. This causes OLS results to be consistent but inefficient. The White correction provides consistent estimates of the standard errors, and it is these that are used in calculating the reported t-values.

Other Results

The results on the impact of union membership are interesting. Introducing variables that measure firm size into the basic equation of R5.1 reduces the coefficient on *union* by 4.6 percentage points in R5.3. This matches almost exactly with the 5 percentage point bias found in Miller and Mulvey (1996). This effect is produced only by the firm size variables and not by the hierarchy variables. In R5.2 the *union* coefficient shows negligible difference to that in R5.1. Including both firm size and hierarchy variables reduces the effect, as compared to the basic model, to only a 2.7 percentage point reduction. However, there is in all four models a significant union membership effect. This is similar to other studies of the same period surveyed in Miller and Mulvey (1996). This is an example of one dimension in which firm size and hierarchy variables do not have the same effect.

Miller and Mulvey (1996) show that for the 1993 Survey of Training and Education the union effect is negligible (after controlling for firm size) in a regression including a rich array of variables on the individual. Miller and Mulvey (1996, p144) also replicate models from other studies with their data set and find a pattern which they claim is basically the same as in the original studies. Their claim is misleading, because although the replications have a similar pattern of relative magnitudes, the actual magnitudes are consistently smaller. In nine out of eleven comparisons Miller and Mulvey's (1996, p142-143) replications produce lower union effects.

The results from Table 5.5 shows that the union effect in the middle 1980's is significantly different to that in 1995. In all four models, the coefficient on *union*$\times$*year95* is sufficiently large and negative as to indicate no positive wage effect from union membership in 1995. This supports Mulvey and Miller's finding of a negligible wage effect for the mid 1990's, but casts doubt on their conclusion that previous findings of a union effect are due solely to omitted variable bias. The results presented here indicate that the presence of a union effect is dependent on the period considered.

Table 5.5

OLS Regression Results
Dependent Variable is *lnwages* (t-stats in parenthesies).

	R5.1	R5.2	R5.3	R5.4
year86	0.060	0.048	0.075	0.067
	(1.868)	(1.544)	(2.460)	(2.225)
year87	0.175	0.175	0.181	0.184
	(5.532)	(5.742)	(5.851)	(6.079)
year95	0.542	0.547	0.559	0.567
	(8.190)	(8.543)	(8.637)	(8.896)
education	0.068	0.061	0.062	0.057
	(11.881)	(10.814)	(11.091)	(10.153)
experience	0.037	0.032	0.035	0.031
	(8.074)	(7.053)	(7.818)	(6.950)
$experience^2/100$	-0.059	-0.052	-0.056	-0.051
	(-6.393)	(-5.796)	(-6.283)	(-5.766)
married	0.057	0.055	0.057	0.054
	(1.751)	(1.691)	(1.842)	(1.735)
metro	0.093	.072	0.071	0.064
	(3.756)	(2.962)	(2.952)	(2.688)
union	0.146	0.144	0.010	.120
	(5.566)	(5.549)	(3.799)	(4.573)
below=1		0.092		0.085
		(3.401)		(3.194)
below=2		0.219		0.180
		(7.279)		(6.003)
above=1		-0.017		-0.017
		(-0.438)		(-0.456)
above=2		0.101		0.026
		(3.057)		(0.756)
married× *year95*	0.117	0.104	0.115	.103
	(1.684)	(1.552)	(1.701)	(1.551)
union× *year95*	-0.161	-0.157	-0.186	-0.173
	(-2.650)	(-2.700)	(-3.188)	(-3.023)
firm size: 10-99			0.151	0 .121
			(3.928)	(3.071)
firm size: 100+			0.272	0.223
			(7.329)	(5.467)
constant	0.913	0.937	0.847	0.905
	(10.969)	(10.962)	(10.619)	(10.713)
N	975	975	975	975
Adjusted-R^2	0.478	0.510	0.513	0.523
Reset ($F[3, N-k-3]$)	3.16	1.50	1.56	0.63
Cook-Weisberg ($\chi^2(1)$)	12.68	8.54	6.37	5.36
White ($\chi^2(1)$)	15.23	10.43	8.28	7.08

5.5 Conclusion

The brief survey of the grand contract literature showed that there are theoretical reasons for thinking that organizational structure, or more particularly hierarchy shape and hierarchical position, can be important in determining wages. Also this theoretical hierarchical effect is not due solely to the sorting of ability.

The empirical findings indicate that both human capital variables and organizational/hierarchical variables are important in determining wages. The most robust result was that wages increase with level in the hierarchy, by approximately 9% per level. Failing to account for this effect overestimates the effect of education and experience by approximately 20%.

It appears that there is a second organizational effect beyond level in the hierarchy. This second, positive effect is determined by the size of the organization or hierarchy. This may be to do with the number of levels above an individual, or may be due to other, so called firm size effects such as a monopsony effect or regulation. The results are somewhat confounded.

The Qian model fits better overall with the empirical results, but the Calvo and Wellisz (1979) model is superior in its consideration of skill/ability and hierarchical position. A combination of the two models might be better than either individually. There are other models of hierarchies not considered, and the models here are far from the final word on hierarchies. However, these models have provided a useful framework for considering an aspect of jobs based analysis within an extension of the traditional human capital approach. The empirical results indicate which of the current models fit best with the real world, and perhaps more importantly, indicate the properties that future models must have.

There is a clear need for further research into the effects of organizational structure on wages. In particular, the firm size effects needs to be disaggregated into a number of effects with discernible causes. The work presented here indicates the importance of organizational structure in wage determination and sheds light on some of the relevant issues by bringing together theoretical work on hierarchies with empirical testing.

5.6 Appendix A

Following the original exposition from Calvo and Wellisz(1979, p120-122) the result is shown by considering two simple hierarchical arrangements. In the first, the owner supervises workers directly (there are only two tiers) in the second case there are three tiers, so that the owner supervises supervisors who in turn supervise the workers.

When there are two tiers

$$p_2 = 1/s_1 \tag{5.2}$$

hence

$$w_2 = (s_1 - 1)\, k + h(\beta_2)\,. \tag{5.3}$$

The profit for the two level firm, π_2, can be written as

$$\pi_2 = \beta_2 x_2 - w_2 x_2 = \beta_2 s_1 - w_2 s_1. \tag{5.4}$$

Substituting in equation 5.3 gives

$$\pi_2 = \beta_2 s_1 - [(s_1 - 1) k + h(\beta_2)] s_1. \tag{5.5}$$

Thus the firms maximization problem is

$$\max_{\beta_2, s_1} \pi_2 \equiv \pi_2^*. \tag{5.6}$$

Assuming the existence of an interior solution gives at the optimum

$$h'(\beta_2) = 1. \tag{5.7}$$

Now consider the three level case in which the owner supervises $x_2 = s_1$ supervisors who supervise $x_3 = s_2 s_1$ production workers. The respective probabilities for detecting someone shirking at level l are

$$p_2 = \frac{1}{x_2} \tag{5.8}$$

and

$$p_3 = \frac{\beta_2}{s_2} = \frac{\beta_2 x_2}{x_3}. \tag{5.9}$$

The respective wages are

$$w_2 = (x_2 - 1) k + h(\beta_2) \tag{5.10}$$

and

$$w_3 = \left(\frac{x_3}{\beta_2 x_2} - 1\right) k + h(\beta_3). \tag{5.11}$$

Total profit for this three level firm, π_3 is the output of the workers less the wages payments for the workers and the level 2 supervisors:

$$\pi_2 = \beta_3 x_3 - w_2 x_2 - w_3 x_3. \tag{5.12}$$

Substituting in the expressions for the wages gives

$$\pi_2 = \beta_3 x_3 - \left[\left(\frac{x_3}{\beta_2 x_2} - 1\right) k + h\left(\beta_3\right)\right] x_3 - \left[\left(x_2 - 1\right) k + h\left(\beta_2\right)\right] x_2. \tag{5.13}$$

By definition

$$\begin{aligned} \pi_2^* &\equiv \max_{\beta_2, x_2, \beta_3, x_3.} \pi_2 \qquad (5.14) \\ &= \max_{\beta_2, x_2} \left(\begin{array}{c} \max_{\beta_3, x_3} \left\{\beta_3 x_3 - \left[\left(\frac{x_3}{\beta_2 x_2} - 1\right) k + h\left(\beta_3\right)\right] x_3\right\} \\ - \left[\left(x_2 - 1\right) k + h\left(\beta_2\right)\right] x_2 \end{array} \right). \end{aligned}$$

For simplicity let $\widetilde{s}_2$ be the effective, as opposed to the physical, span of control at level 2, then

$$\widetilde{s}_2 = \frac{x_3}{\beta_2 x_2}.$$

Substituting and rearranging:

$$\pi_2^* = \max_{\beta_2, x_2} \left(\begin{array}{c} \beta_2 x_2 \max_{\beta_3, \widetilde{s}_2} \left\{\beta_3 \widetilde{s}_2 - \left[\left(\widetilde{s}_2 - 1\right) k + h\left(\beta_3\right)\right] \widetilde{s}_2\right\} \\ - \left[\left(x_2 - 1\right) k + h\left(\beta_2\right)\right] x_2 \end{array} \right). \tag{5.15}$$

Thus from equations it follows that

$$\pi_2^* = \max_{\beta_2, x_2} \left(\beta_2 x_2 \pi_1^* - \left[\left(x_2 - 1\right) k + h\left(\beta_2\right)\right] x_2\right). \tag{5.16}$$

5.7 Appendix B: Derivation of an Earnings Function

The following derivation of an earnings function is based on pioneering work in Mincer (1974) and Mincer(1993, pp74-75) and a standard contemporary exposition of a Mincer style empirical earnings function from Chapman and Iredale(1993, Appendix A).

Let E_t be gross earnings in period t and C_t is the dollar amount of net investments made in period t. Then if r_t is the rate of return on net investment made in period t

$$E_t = E_{t-1} + r_{t-1}C_{t-1}, \tag{5.17}$$

since any increase in earnings at t over those at $t-1$ can only be the result of a new income stream arising from new investment. In general, r_{t-1} could differ between time periods and between individuals, we however suppress the subscript and use r the average rate of return on the individuals investments in human capital.

The dollar amounts of investments in human capital are not readily observable. Instead, if investment is included as time spent in investment, then the ratio of investment expenditure to gross earnings can be used as time-equivalent units of investment.

Let

$$k_t = C_t/E_t \tag{5.18}$$

then

$$E_t = E_{t-1}(1 + rk_{t-1}) \tag{5.19}$$

and by recursion

$$E_t = E_0(1 + rk_0)(1 + rk_1)...(1 + rk_{t-1}). \tag{5.20}$$

Now for small values of $rk_i, \ln(1 + rk_i)$ can be approximated by rk_i. Thus equation 5.20 can be written:

$$\ln E_t = \ln E_0 + r\sum_{i=0}^{t-1} k_i. \tag{5.21}$$

Analyzing the schooling and post schooling experiences individually, the sum can be separated to run over the years of schooling (up to year s) and the post school years, giving

$$\ln E_t = \ln E_0 + r_s\sum_{i=0}^{s-1} k_i + r_p\sum_{j=s}^{t-1} k_j. \tag{5.22}$$

The assumption on rates of return is relaxed to allow different returns between investment in school and post school investment (denoted r_s and r_p, respectively). Now k_i is taken as not being that different from 1 during the school years. Thus (2.6) becomes

$$\ln E_t = \ln E_0 + r_sS + r_p\sum_{j=s}^{t-1} k_j. \tag{5.23}$$

Post schooling investments are expected to decline over the lifetime (as retirement approaches the expected returns from new investment falls). Hence equation 5.23 may be approximated by including a quadratic experience term. Thus the standard wage equation to estimate is:

$$\ln W_i = a + bS_i + cEXP_i + dEXP_i^2 + eZ_i + u_i. \tag{5.24}$$

Where i denotes these are the characteristics of individual i, $\ln W$ is the natural logarithm of hourly wages, S is years is schooling, EXP is length of time in the labor force, Z is a vector of other wage determining variables and u is a randomly distributed error term.

5.8 Appendix C: Diagnostic Tests

The results of tests for mis-specification and heteroscedasticity are reported at the bottom of Table 5.5. The form of the tests is now briefly described for a regression of the form $Y = XB + \varepsilon$.

The Reset test, is a test for mis-specification in the regression equation. The version here performs the regression $Y = XB + ZT + \varepsilon$, where $Z = \left\{\hat{Y}^2, \hat{Y}^3, \hat{Y}^4\right\}$, and then performs an F test to determine if $T = 0$. Where the null hypothesis ($T = 0$) is no mis-specification.

Only the simple regression of R5.1 is mis-specified. Including either the firm size or hierarchy variables eliminates this mis-specification, again indicating the importance of organizational structure in determining wages.

Cook and Weisberg is a test for heteroscedasticity. It amounts to testing

$T = 0$ in the regression $\text{var}[\varepsilon] = \sigma^2 \exp\left(\hat{Y}T\right)$. See Cook and Weisberg(1983) for details. White is the standard White test for heteroscedasticity. It tests $T = 0$ in the regression of $\text{var}[\varepsilon]$ on the fitted values of the dependent variable. In both cases the null is homoscedasticity.

All the models exhibit heteroscedasticity. This causes OLS results to be consistent but inefficient. The White correction provides consistent estimates of the standard errors, and it is these that are used in calculating the report values.

Chapter 6

Conclusion

6.1 The Production of Management

This thesis has analyzed management as a production process that occurs across a structure of individuals. Two general classes of models have been used to describe this process: information processing models and supervisions models. The differences between the two classes of models are most easily characterized by considering the direction in which the management product flows.

In the information processing approach, there is a decision to be made on the basis of some information. Information is split amongst a number of managers so that different parts of the decision can be worked on concurrently, decreasing the elapsed time taken to make the decision. Each manager works on their part of the problem and then passes their result/conclusion up to a superior. In this way, output flows up the hierarchy, finally merging into one result, which is the decision.

The flow is in the opposite direction in the supervision models. Managerial effectiveness and monitoring work their way recursively down the hierarchy. Total output for a firm is determined by the effort of production workers at the bottom of the hierarchy, and by the effectiveness with which the production workers are managed. Monitoring to ensure effort, and managerial effectiveness are both applied to a production worker by her immediate supervisor. In a similar way the output (of supervision and management) by the supervisor are regulated by the monitoring and supervision supplied by her superior. In this way these twin factors of monitoring and management flow down the hierarchy, from the owner to the production workers, to determine total output.

Three things were done with these models in the proceeding chapters. In chapter 2, some extensions to the basic information processing model were considered. In chapters 3 and 4, a real time version of the information processing model was applied to the problem of production location within a differentiated product model from industrial organization. Chapter 5 is an empirical analysis of the effect of hierarchies on wages, with particular attention paid to testing the varying conclusions from different versions of the grand contract supervision model.

6.2 Main Findings

Chapter 2 considered two extensions to the basic one shot model of information processing proposed by Radner(1993). Those extensions are time based measures for the managerial labor input and fallible performance by managers.

The original information processing model used the number of managers to measure the managerial labor input. Clearly this measure fails to take into account the connection normally made between the amount of time a person works and their pay. The piece rate regime just pays for the time people actually work. The alternative is that people are also paid while they are idle between tasks. The salary regime is proposed as an approximation for this in the one shot model.

It is shown, somewhat surprisingly, that the set of efficient hierarchies is equivalent under these different measures of managerial resource cost. The returns to scale with a linear loss function are also considered for the salary regime and show results different to the fix cost per employee situation considered in Radner (1993).

The second half of chapter 2 proposed two ways in which managers can become fallible. They can make mistakes in the result they pass up the hierarchy or they can be late in passing the message. Here these two types of fallibility are used to more finely characterize the choice of hierarchy by a firm. More generally these concepts are important for developing a theory of fallible networks, which would have immediate application to areas such as banking, nuclear power management and air traffic control, where a clear bad or catastrophic decision exists.

Chapters 3 and 4 contain the major theoretical contribution of the thesis. They show how the relationship between hierarchical structure and product market behavior can be explicitly modeled. This synthesis between traditional industrial organization and formal organizational theory appears to offer an alternative to the behavioral based theories of the firm which currently form the bridge between economics and management.

The model considers a real time decision problem with endogenous choice of the quantity of information. Specifically it was shown that the optimal choice of organization for a monopolist in a dynamic differentiated products market is determined by the speed at which market conditions change.

The model assumed that consumer preferences changed over time, according to a stochastic process and were not observable. The analysis was initially applied to a fixed price settling and then extended to allow price to be set optimally once location was chosen. Simulations showed that profit was increasing in the width of appeal for the product and decreasing in the speed of change in market conditions. A product which can capture a larger market at the same cost will in general produce a larger profit. As the shocks to consumer preferences increase the accuracy of any forecast will decrease and hence profits will also decrease.

The dependence of optimal organizational size on market conditions was also shown. Taking a larger market research sample requires a larger organization and gives a good picture of the past. A smaller organization using less data can act faster but has a poorer historical picture. Thus there are forces pushing for both a larger and a smaller firm. The relative magnitudes of these two forces is determined by the rate at which consumer preferences change, which therefore determines optimal firm size.

The formal model gives other interesting insights. For example, suboptimal choice of hierarchy is less critical for both stable environments and products with smaller widths of appeal. It is also shown that specialized products attain a higher expected fraction of their potential maximum profit than general products do.

Chapter 5 examines the other market on which a firm's hierarchy has an impact: the labor market. The supervisory models of hierarchies are currently better suited to addressing questions of wage determination than the information processing models which were used in the earlier chapters. No new theoretical results are proved in this chapter, rather the existing theory is surveyed and interpreted so as to produce empirically testable conclusions. The variants on the grand contract supervisory model all predict that wages will increase with level in the hierarchy. The implications for the effect of the number of levels above an individual covered all possibilities, positive, negative and zero.

Empirical analysis showed that wages did indeed increase with level in the hierarchy, and that this result is robust to various model specification. It is also shown that omitting the hierarchy variables from a traditional wages equation overestimates the effects of education and experience by about 10%. The other results show that beside level in the hierarchy, the structure of the firm is also important. However, it is not possible to tell if it is the increasing size of the hierarchy or the increasing firm size that have a positive affect on wages.

The thesis highlights a number of important areas for future research, the most obvious is the extension of the monopolist hierarchy analysis to an oligopoly setting. This is a subject of current research. The empirical section provides some important new facts for those modeling hierarchies, but also indicates the need for more facts. This is also a current topic of research with case studies on hierarchical structure in a number of large firms.

Bibliography

Baker, G. M. Gibbs and B. Holmstrom (1994a), "The Internal Economics of the Firm: Evidence from Personnel Data", Quarterly Journal of Economics, November 1994, 109(4), 881-919.

Baker, G. M. Gibbs and B. Holmstrom (1994b), "The Wage Policy of a Firm", Quarterly Journal of Economics, November 1994, 109(4), 921-955.

Baker, G. and B. Holmstrom (1995), "Internal Labor Markets: Too Many Theories, Too Few Facts", American Economic Review, 85, 2, 255-259.

Baumol W., S. Blackman and E. Wolfe (1989), *Productivity and American Leadership: The Long View*, Chapter 7, Cambridge, MA , MIT Press.

Bean, Clive S. (1991), "Comparison of National Social Science Survey Data with the 1986 Census." National Social Science Survey Report 2(6):12-19 ISSN 1031-4067.

Beckmann(1960), "Some Aspects of Returns to Scale in Business Administration", Quarterly Journal of Economics, LXXIV(1960), 257-65.

Beckmann, M.(1977), "Management Production Function and the Theory of the Firm", Journal of Economic Theory, 14, 1-18.

Bolton, P. and M. Dewatripont(1994), "The Firm as a Communication Network", Quarterly Journal of Economics, 109, 809-839.

Calvo, G. and S. Wellisz (1978), "Supervision, Loss of Control and the Optimal Size of the Firm", Journal of Political Economy, 86, 943-952.

Calvo, G. and S. Wellisz (1979), "Hierarchy, Ability, and Income Distribution", Journal of Political Economy, 87, 115-134.

Carroll, P. (1993), *Big blues : the unmaking of IBM*, New York,Crown Publishers.

Carroll, G and M.T. Hannan(1989), *Organizational Ecology*, Cambridge, MA, Harvard University Press.

Cook, R.D and S. Weisberg (1983), "Diagnostics for heteroscedasticity in regression", Biometrika (1983), 70, 1, 1-10.

Cusumano, M.A. (1985), *The Japanese Automobile Industry: technology and management at Nissan and Toyota*, Cambridge, MA, Harvard University Press.

Cyert, R.M. and J.G. March (1963), *A Behavioral Theory of the Firm*, Englewood Cliffs, NJ., Prentice-Hall.

Gittleman, M. and E. N. Wolff (1995), "R&D activity and cross-country comparisons", Cambridge Journal of Economics, No. 19, pp189-207.

Gordon, D. (1990), "Who Bosses Whom? The Intensity of Supervision and the Discipline of Labor", American Economic Review, 80, 2, 28-32.

Harrington, J.E., Jr (1995), "Experimentation and Learning in a Differenetiated-Product Duopoly", Journal of Economic Theory, 66, 275-288.

Kelley, Jonathan, Clive S.Bean and M.D.R.Evans. (1988). Australia, 1987: National Social Science Survey. Codebook and Machine Readable Data File. Canberra: International Survey Center, Institute of Advanced Studies, Australian National University.

Kelley, Jonathan, Robert G. Cushing and Bruce Headey(1984). AUSTRALIAN NATIONAL SOCIAL SCIENCE SURVEY, 1984. Codebook and Machine Readable Data File. Canberra: International Survey Center, Institute of Advanced Studies, Australian National University.

Kelley, Jonathan, M.D.R.Evans and Clive S.Bean. (1995). Australia, 1989 - 1993: International Social Science Survey – Australian Family Panel. Codebook and Machine Readable Data File. Canberra: International Survey Center, Institute of Advanced Studies, Australian National University.

Kelley, Jonathan, M.D.R.Evans and Clive S.Bean. (1987). Australia, 1986: National Social Science Survey. Codebook and Machine Readable Data

File. Canberra: International Survey Center, Institute of Advanced Studies, Australian National University.

Keren, M. and D. Levhari (1979), "The optimal span of control in a pure hierarchy", Management Science, 11, 1162-1172.

Keren, M. and D. Levhari (1983), "The internal organization of the firm and the shape of average costs", The Bell Journal of Economics, 14, 474-486.

Keren, M. and D. Levhari (1989), "Decentralization, aggregation, control loss and costs in a hierarchical model of the firm", Journal of Economic Behavior and Organization, 11, 213-236.

Koza, John R. (1992), *Genetic Programming*, A Bradford Book, MIT, 1992.

Lazear, E.P (1992),"Compensation, Productivity and the New Economics of Personnel", Hoover Institute Working Papers in Economics: E-92-9, April 1992.

Lazear, E.P (1995), "A Jobs-Based Analysis of Labor Markets", American Economic Review, 85, 2, 260-265.

Marschak, J. and R. Radner (1972), *Economic Theory of Teams*, New Haven, Yale University Press.

Meagher, K. (1996a), "Efficient Hierarchies: Equivalence under differing employment regimes", Proceedings of 1996 Econometric Society Australasian meeting.

Meagher, K. (1996b),"Efficient Hierarchies: Equivalence under differing employment regimes", ANU Working Papers in Economics and Econometrics No. 309.

Meagher, K. (1996c), "Managing Change and the Success of Niche Products", Santa Fe Institute Working Paper 96-08-066.

Meagher, K. and D. Miron (1997), "Real-time Decentralized Bayesian Decision Making", Australian National University.

Miller, P. and C. Mulvey (1996), "Unions, Firm Size and Wages", The Economic Record, 72, 217, June 1996, 138-153.

Mincer, J. (1974), *Schooling, experience, and earnings*, New York, National Bureau of Economic Research.

Mincer, J. (1993), *The collected essays of Jacob Mincer*, Aldershot, England, Edward Elgar.

Oi, W.Y. (1988), "Are Workers Overpaid By Big Firms?", Australian Journal of Management, 13, 2, December 1988, 203-221.

Osborne, M.J. and A. Rubinstein (1994), *A Course In Game Theory*, The MIT Press, Cambridge, Massachusetts.

Peters and Waterman (1982), *In Search of Excellence*, Harper & Row Publishers Inc., New York.

Preston, A. (1997), "Where Are We Now With Human Capital Theory in Australia?", The Economic Record, 73, 220, March 1997, 51-78.

Qian, Y. (1994), "Incentives and Loss of Control in an Optimal Hierarchy", Review of Economic Studies, 61, 527-544.

Radner, R.(1992), "Hierarchy: The Economics of Managing", Journal of Economic Literature, 30 (1992), 1382-1414.

Radner, R. (1993), "The Organization of Decentralized Information Processing", Econometrica, Vol. 61, No. 5 (September, 1993), 1109-1146.

Radner, R. and T. Van Zandt (1992), "Information Processing in Firms and Returns to Scale," Annales d'Economie et de Statistique, No. 25-26, pp265-298.

Radner, R. (1996), "Bounded Rationality, Indeterminacy, and the Theory of the Firm ", Economic-Journal, 106(438), September 1996, pages 1360-73.

Rosen, S. (1982), "Authority, control and the distribution of earnings", Bell Journal of Economics, Vol. 13 (1982), 311-323.

Sah R. and J. Stiglitz (1986), "The Architecture of Economic Systems: Hierarchies and Polyarchies," American Economic Review, LXXVI , pp716-727.

Simon, H. (1953), *Administrative Behavior*; The Macmillan Co., New York, 1953.

Smith, A. (1982), *The Wealth of Nations*, Penguin Books Ltd., New York, NY, USA. First published in 1776.

Sobel (1992), "How to Count to One Thousand", The Economic Journal, 102 (January 1992), pp1-8.

Spanos, A.(1986), *Statistical foundations of econometric modelling*, Cambridge University Press.

Tirole, J. (1988), *The Theory of Industrial Organization*, Cambridge, MA, MIT Press.

Van Zandt, T. (1995a), "The structure and returns to scale of organizations that process information with endogenous agents", unpublished working paper, Princeton University.

Van Zandt, T. (1995b), "Continuous approximations in the study of hierarchies", Rand Journal of Economics, 26, 575-590.

Van Zandt, T. (1997a), "Decentralized information processing in the theory of organizations", in M. Sertel (Ed.), *Contemporary Economic Developments Reviewed, Vol. 4: The Enterprise and its Environment,* London, MacMillian Press (forthcoming).

Van Zandt, T. (1997b), "Organizations with an endogenous number of information processing agents", in M. Majumdar (Ed.), *Organizations with Incomplete Information,* Cambridge, Cambridge University Press (forthcoming).

Van Zandt, T. (1997c), "The Scheduling and Organization of Periodic Associative Computation", Economic Design (forthcoming).

Van Zandt, T. (1997d), "Real-time hierarchical resource allocation", Princeton University.

Van Zandt, T. (1997e), "Real-time hierarchical resource allocation with quadratic costs", Princeton University.

Van Zandt, T. (1997f), "Structure and returns to scale of real-time hierarchical resource allocation", Princeton University.

Van Zandt, T. (1997g), "Real-time decentralized information processing as a model of organizations with boundedly rational agents", Review of Economic Studies (forthcoming).

Van Zandt, T. and R. Radner(1997), "Real-time decentralized information processing and returns to scale", unpublished working paper, Princeton University.

Williamson, O. (1967), "Hierarchical control and optimum firm size", The Journal of Political Economy, 75, 2 (April 1967), 123-138.